JJMIE

Volume 13, Number 4, December. 2019
ISSN 1995-6665
Pages 247 - 251

Jordan Journal of Mechanical and Industrial Engineering

A New High Accuracy Mathematical Approximation to the Cumulative Normal Density Function

[a]Atif Alkhazali, [b]Mohammad Al-Rabayah, [a]Mohammad M. Hamasha

[a]Department of Industrial Engineering, The Hashemite University, Az Zarqa 13133, Jordan
[b]College of Electrical Engineering, American University of the Middle East, Eqaila, Kuwait

Received 1 July. 2019

Abstract

This paper develops two approximations to the normal cumulative distribution function. Although the literature is rich in approximation functions for the normal distribution, they are not very accurate. This proposed approximation is an enhanced logistic cumulative function approximation to the normal cumulative distribution function. The paper starts with approximating the logistic function to the normal distribution and then introduces a second term to the first approximation to increase the accuracy. Besides the simplicity of the introduced model, it has a maximum error of less than 0.0012 for the entire range. This level of accuracy is superior if compared to other introduced models by other authors. The current can be used to estimate the normal distribution based probabilities and associated statistics. The deviation of the proposed model from the actual normal cumulative distribution with Z score is discussed at the end of this paper.

Keywords: Normal distribution, logistic function, normal cumulative distribution, approximation;

1. Introduction

The normal distribution is one of the most important continuous probability distribution functions used in the field of engineering and science. This is due to the fact that many natural phenomena can be accurately described using this distribution, and that the error measurements follow the normal distribution in theory. Further, the normal distribution is important in predicting solutions in many engineering fields such as heat flow, operations research, mechanics, quality engineering, and reliability engineering.

The normal distribution density function with mean of μ and standard deviation of σ is addressed in equation (1)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

The normal distribution density function with μ=0 and σ=1 is referred by standard normal distribution as addressed in equation (2)

$$\emptyset\,(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{2}$$

The normal distribution can be transferred into standard normal distribution to be handled by the popular z-table using the transformation formula (i.e., Equation 3).

$$z = \frac{x-\mu}{\sigma} \tag{3}$$

In statistics, the distribution cumulative density function is the probability that the selected random value is equal or less than specific value *c*, or in other words, P(X<c). For the normal distribution, the cumulative distribution function is a complex integral function and usually denoted with the capital Greek letter, Φ as in the Equation (4).

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt \tag{4}$$

Unfortunately, the above integral cannot be solved into a closed-form formula, and the results of this integral are usually found using calculators, Excel spreadsheets, as well as the popular z- table. One practical problem arises when using such tables, is that those tables only provide the cumulative probabilities for certain discrete values of z. In many scenarios, the required z value will not be available on the table, and therefore, the value will be approximated using adjacent z values.

In order to overcome this practical problem, many researchers have proposed a number of approximate functions for the cumulative normal distribution. However, many proposed approximate functions are either mathematically complicated and lack accuracy, or not valid for the entire range of z values. This paper proposes an enhanced logistic cumulative function approximation to the normal cumulative distribution function. Besides the simplicity of the introduced model, it has a maximum error of less than 0.0012 for the whole range, [-∞:∞].

This paper is organized as follows: In Section 2, a literature review on similar approximate functions with

* Corresponding author e-mail: atif@hu.edu.jo

the introduced one is conducted. Section 3 discusses our proposed model and shows our findings, and finally, section 4 provides a short conclusion on our work.

2. Review of Previous Work

Many proposed approximation functions of normal density are introduced in the literature. However, we can gather these approximations into two main types: approximation based on numerical algorithms and Ad-hoc approximations [1]. This section presents a historical review of different proposed approximations of the cumulative normal distribution function. The results of this review are summarized in Table 1.

Table 1. Summary of approximate functions of the cumulative normal distribution.

Author(s)	Year	Approximate Function
Yerukala and Boiroju [2]	2015	$\Phi(z) \approx 1-\frac{\exp(-z^2/2)}{\left(\frac{44}{79}+\frac{8}{5}z+\frac{5}{6}\sqrt{z^2+3}\right)}$
Choudhury [3]	2014	$\Phi(z)\approx 1-\frac{1}{\sqrt{2\pi}}\frac{e^{\frac{-z^2}{2}}}{0.226+0.64z+0.33\sqrt{z^2+3}}$
Yerukala [4]	2012	$\Phi(z)\approx\begin{cases}0.46375418+0.065687194H_{11}-0.602383931H_{12} & ;0\le z\le 3.6\\ 1 & ;z>3.6\end{cases}$
Aludaat and Alodat [5]	2008	$\Phi(z)\approx 0.5+0.5\sqrt{1-e^{-\sqrt{\frac{\pi}{8}}z^2}}$
Bryc [6]	2002	$\Phi(z)\approx\frac{z^2+a_1z+a_2}{\sqrt{2\pi}z^3+b_1z^2+b_2z+2a_2}e^{\frac{-z^2}{2}}$
Waissi and Rossin [7]	1996	$\Phi(z)\approx\frac{1}{1+\exp\left(-\sqrt{\pi}\left(0.9z+0.0418198z^3-0.0004406z^5\right)\right)}$
Lin [8]	1990	$\Phi(z)\approx 1-\frac{1}{1+e^{y}}$
Norton [9]	1989	$\Phi(z)\approx\begin{cases}1-0.5e^{-\left(\frac{z^2+1.2z^{0.8}}{2}\right)} & ;0\le z\le 2.7\\ \frac{1}{\sqrt{2\pi}z}e^{-\frac{z^2}{2}} & ;z>2.7\end{cases}$
Lin [10]	1989	$\Phi(z)\approx 1-0.5\left(e^{-0.717z-0.416z^2}\right)$
Hammakar [11]	1978	$\Phi(z)\approx 1-0.5\left\{1-\left(1-e^{-y^2}\right)^{0.5}\right\}$
Hart [12]	1966	$\Phi(z)\approx 1-\frac{e^{\frac{-z^2}{2}}}{\sqrt{2\pi}z}\left(1-\frac{\frac{\sqrt{1+bz^2}}{1+az^2}}{P_0z+\sqrt{P_0^2z^2+e^{\frac{-z^2}{2}}\frac{\sqrt{1+bz^2}}{1+az^2}}}\right)$
Zelen and Severo [13]	1964	$\Phi(z)\approx 1-\left(a_1t-a_2t^2+a_3t^3\right)\left(\frac{e^{\frac{-z^2}{2}}}{\sqrt{2\pi}}\right)$
Tocher [14]	1963	$\Phi(z)\approx\frac{e^{2kz}}{1+e^{2kz}}$, where $k=\sqrt{\frac{2}{\pi}}$
Hart [15]	1957	$\Phi(z)\approx\frac{1}{\sqrt{2\pi}}\left\{\frac{e^{\frac{-z^2}{\pi}}}{z+0.8e^{-0.4z}}\right\}$
Cadwell [16]	1951	$\Phi(z)=1/2\left[1+\left\{1-\exp\left(-\frac{2z^2}{\pi}-\frac{2(\pi-3)z^4}{3\pi^2}\right)\right\}^{1/2}\right]$
Polya [17]	1945	$\Phi(z)\approx\frac{1}{2}\left\{1+\left(1-e^{\frac{-2z^2}{\pi}}\right)^{\frac{1}{2}}\right\}$

Most of the above-mentioned models have an accuracy range from 0.01 to 0.001.

3. Proposed Model

The proposed model uses the logistic function to build the first term of our approximation. The logistic function is a continuous probability density function that shares the normal distribution in the shape (i.e., bell shape with symmetry about the mean). However, it has heavier tails (higher kurtosis). Equations (5) and (6) represent the logistic density and cumulative logistic functions, respectively.

$$f(z) = \frac{e^{-\frac{z-\alpha}{\beta}}}{\beta\left(1+e^{-\frac{z-\alpha}{\beta}}\right)^2}, -\infty < z < \infty \quad (5)$$

$$F(z) = \frac{1}{1+e^{-\frac{z-\alpha}{\beta}}}, -\infty < z < \infty \quad (6)$$

The best cumulative logistic approximation to the cumulative normal distribution can be achieved with using the ratio α/β of 1.702. See Equation (7). Figure 1 shows the deviation between the logistic cumulative function and the cumulative normal distribution of at α/β = 1.702. In spite of the best approximation we achieved, the error is significant and cannot be used as a good approximate without modification. The best logistic approximation to the normal distribution is shown in Equation (8). These introduced approximations are very simple compared with any other approximations, and the accuracy level is very acceptable for most engineering and science applications.

$$\Phi_1(z) \approx \frac{1}{1+e^{-1.702z}}, -\infty < z < \infty \quad (7)$$

$$\emptyset(z) \approx \frac{1.702e^{-1.702z}}{(1+e^{-1.702z})^2}, -\infty < z < \infty \quad (8)$$

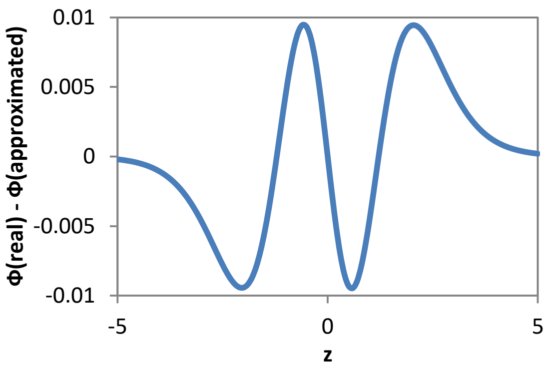

Figure 1. The deviation of best fit logistic cumulative function to the cumulative normal distribution.

The previous approximation can be enhanced to be a more accurate approximation. We noticed that the shape of deviation vs. z curve can be approximated in the middle with a damped negative sign wave. with discarding the value of sign wave (deviation) over the range of -2 to 2, we significantly reduce, as addressed in Equation (9).

$$\Phi_2(z) = \frac{1}{1+e^{-1.702z}} - 0.0095\sin(2.5z)\,e^{-0.01z^2} \quad -\infty \leq z \leq \infty \quad (9)$$

Although the approximation is very good at -2<z<2, the deviation is high outside of this region, as shown in Figure 2. Therefore, we developed a separate second term to be discarded from the best-fit logistic function for the region (-∞<z<-2 and 2<z<∞). The final approximation is addressed in Equation (10).

$$\Phi_3(z) \cong \begin{cases} \dfrac{1}{1+e^{-1.702z}} - 0.0095\sin(2.5z)\,e^{-0.01z^2}, |z| \leq 2 \\ \dfrac{1}{1+e^{-1.702z}} + 10^{-3}(0.2656|z|^5 - 5.197|z|^4 + 39.81|z|^3 - 147.6|z|^2 + 258.64|z| - 163.3), |z| \geq 2 \end{cases} \quad (10)$$

The deviation between the introduced model results and real standard normal cumulative distribution results with Z score is shown in Figure3. The maximum absolute deviation is 0.0012 at z=1.1 and z=-1.1.

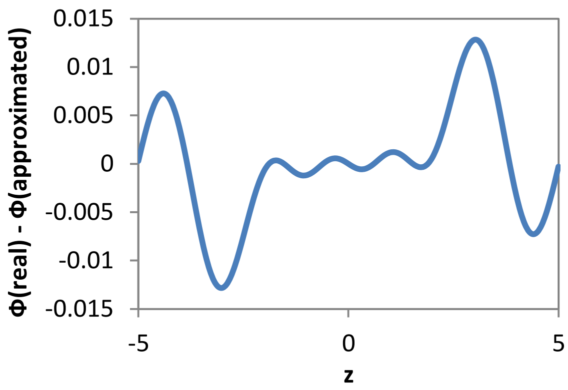

Figure 2. The deviation of Equation (9) cumulative density function approximation from the actual values.

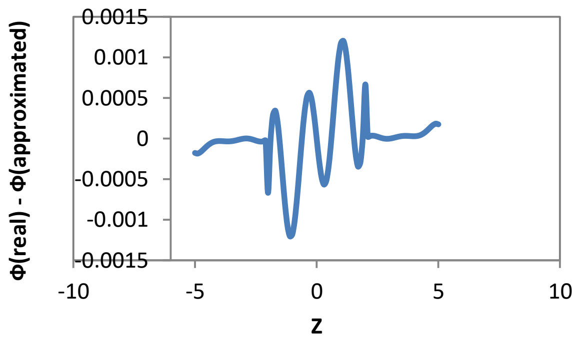

Figure 3. The deviation of Equation (10) cumulative density function approximation from the actual values.

The proposed approximate function is competitive compared to the cumulative logistic approximation function, as well as to other different proposed approximation functions. Such comparison is shown in Table 2.

Table 2. Illustration of the proposed model accuracy compared to cumulative logistic function and other approximation functions accuracies in term of the deviation approximation from the actual values.

Z	Φ(Introduced)	Logistic	Cadwell	Lin 1988	Lin 1989	Lin 1990	Bryc
-4	2.96098E-05	0.00107201	3.08567E-05	3.06079E-05	4.86714E-06	5.62865E-06	1.03553E-05
-3.8	3.24454E-05	0.001478192	6.8885E-05	6.84348E-05	8.35073E-06	7.42694E-06	2.77357E-05
-3.6	3.37333E-05	0.002018822	0.000146014	0.000145852	1.3289E-05	7.86377E-06	6.97113E-05
-3.4	2.65647E-05	0.002721472	0.000292592	0.000295512	1.93103E-05	5.26687E-06	0.000164983
-3.2	1.24453E-05	0.00360615	0.000551839	0.000567549	2.48944E-05	1.69102E-06	0.000368331
-3	7.1871E-08	0.004673872	0.000975503	0.001030156	2.66901E-05	1.22292E-05	0.000776305
-2.8	4.28237E-07	0.005890706	0.001610124	0.001761942	1.91451E-05	2.09886E-05	0.001544627
-2.6	1.39324E-05	0.007169	0.002473075	0.002831687	4.7625E-06	1.61284E-05	0.002899789
-2.4	3.23402E-05	0.008350522	0.003524086	0.004265179	5.05495E-05	1.93149E-05	0.005131209
-2.2	3.38921E-05	0.009199975	0.004645573	0.006006956	0.000115869	9.90841E-05	0.00854618
-2	0.000667975	0.009420556	0.005647979	0.007893744	0.000180436	0.000215536	0.013373235
-1.8	0.000286853	0.008703622	0.006309427	0.009659229	0.000193979	0.000313553	0.019615519
-1.6	0.000189019	0.006818886	0.006442876	0.010981261	6.73163E-05	0.000267983	0.026884312
-1.4	0.000469301	0.003737062	0.005967099	0.011564025	0.000324296	0.000112152	0.034274623
-1.2	0.001074377	0.000247096	0.004950418	0.011228747	0.001113512	0.001024159	0.040363059
-1	0.001177894	0.00445102	0.003603905	0.009978327	0.00237755	0.002543438	0.04339849
-0.8	0.000708202	0.007875015	0.002220613	0.008009102	0.004034403	0.004457013	0.041718158
-0.6	1.72474E-05	0.009459397	0.001080216	0.005663014	0.005707264	0.006146993	0.03437842
-0.4	0.00051429	0.008495484	0.000354031	0.003336229	0.006583897	0.006635908	0.021986536
-0.2	0.000475254	0.005027975	4.69985E-05	0.001375983	0.005314561	0.004846699	0.007844134
0	0	0	0	0	0	0	9.39381E-09
Maximum Absolute error	0.001177894	0.009459397	0.006442876	0.011564025	0.006583897	0.006635908	0.04339849

4. Potential Application of the Model

The normal distribution can describe the probability (uncertainty) of many surrounding of engineering measurements. It, for example, describes the uncertainty of barometric pressure. If the engineer has not the proper device to measure the pressure, he/she can guess center of a normal distribution from the previous data and he/she would know estimate the pressure. With the current model, he/she can approximate the result mathematically with a very close result. The deviation of the model result approximation from the actual values is not noticeable for all engineering application. Furthermore, the normal distribution fits many human performance variations. For example, the commonly IQ is normally distributed and can estimated using the current model. The central limit theorem puts the normal curve as most important continuous distribution. This is besides the fact that many data are very accurately described with normal distribution.

The normal distribution is very popular in the research. In this regard, we find more than 6.5 million published papers on google scholar dealing with normal distribution in various fields. Further, there are 13 papers in Jordan journal of mechanical and industrial engineering handling the normal distribution [18-30].

As practical example, our model will be applied on a selected case study from Jordan Journal of Mechanical and Industrial Engineering. Aljebory and Alshebeb [19] discussed a case study of statistical process control of a chemical product. Figure 4 is a snap shot of $\bar{X}$ control chart for the pH of that product. Assuming a shift happen in the Phase II from $\bar{X} = 9.26$ to $\bar{X} = 10$, we can expect the average run length to detect the shift or, in other words, the number of subgroups required to detect the shift, as follows.

$$3\sigma_{\bar{X}} = 10.278 - 9.26 = 1.018$$

$$\sigma_{\bar{X}} = 0.339$$

$$Shift = 10 - 9.26 = 0.74$$

$$Shift\ in\ term\ of\ \sigma_{\bar{X}} = \frac{0.74}{0.339} = 2.18$$

$$\beta = \Phi(\mathrm{L} - \mathrm{K}) - \Phi(-\mathrm{L} - \mathrm{k}) \cong \Phi(0.82)$$

By using our model, Φ(0.82)=0.79312, which is very close to the actual number 0.7938

Furthermore, the average run length to detect the shift (ARL_1) equals to1/(1- β)=0.79 using the model and it is almost the same actual value.

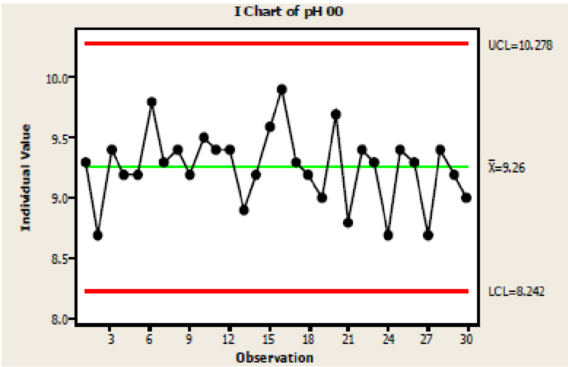

Figure 4. $\bar{X}$ control chart for a chemical product discussed in [19]

The normal distribution occupied a wide area of probability and statistics science. As well known, the probability and statistics is a major area in the industrial and systems engineering, and indeed, it is the background of other major areas of industrial and systems engineering.

5. Conclusion

In this paper, an approximation to the cumulative distribution function is proposed. The best approximation of logistic function to the normal distribution CDF is found. Then, a second part is added to the logistic function approximation to increase the accuracy to reach better accuracy level compared with other introduced models. A numerical comparison shows that our approximation is very accurate with a maximum absolute error of 0.0012 for the entire region.

References

[1] Matic, R. Radoici, D."Stefanica, A sharp Pólya-based approximation to the normal cumulative distribution function", Applied Mathematics and Computation,Vol. 322, 2018, 111–122.

[2] R. Yerukala, N. Boiroju,"Approximation to standard normal distribution function, International journal of Scientific and engineering research", International Journal of Scientific and Engineering Research, Vol. 6, 2015, 515-518.

[3] Choudhury,"A simple approximation to the area under standard normal curve", Mathematical statistics,Vol. 2, 2014, 147-149.

[4] R. Yerukala, "Functional approximation using neural networks", Thesis, Department of Statistics. Osmania University, Hyderabad, Telangana, India.2012.

[5] K.M. Aludaat, M.T. Alodat,"A note on approximating the normal distribution function", Applied Mathematical Science,Vol. 2, 2008, 425-429.

[6] W. Bryc, "A uniform approximation to the right normal tail integral", Applied Mathematics and Computation,Vol. 127, 2002, 365-374.

[7] G.R. Waissi, D.F. Rossin, "A sigmoid approximation of the standard normal integral",Applied Mathematics and Computation,Vol. 7, 1996, 91-95.

[8] J.T. Lin,"A simpler logistic approximation to the normal tail probability and its inverse",Journal of the Royal Statistical Society, Series C,Vol. 39, 1990, 255-257.

[9] R.M. Norton,"Pocket-calculator approximation for area under the standard normal curve", American Statistician,Vol. 43, 1989, 0-1.

[10] J.T. Lin,"Approximating the normal tail probability and its inverse for use on a pocket-calculator",Journal of the Royal Statistical Society, Series C, 38, 1989, 69-70.

[11] H.C. Hammakar,"Approximating the cumulative normal distribution and its inverse",Applied Statistics, Vol. 27, 1978, 76-77.

[12] R.G. Hart,"A close approximation related to the error function (in Technical Notes and Short Paper)", Mathematical Computation, Vol. 20, 1966, 600-602.

[13] M. Zelen, N.C. Severo,"Probability function in handbook of mathematical functions",Dover, New York, pp. 925-995, 1981.

[14] K.D. Tocher,"The art of simulation". English University Press. London. 1963.

[15] R.G. Hart,"A formula for the approximation of definite integrals of the normal distribution function", Tables and other Aids to Computation, Vol. 11, 1957, 265-265.

[16] J.H. Cadwell"The Bivariate Normal Integral", Biometrika, 38, 1951, 475-479.

[17] G. Polya,"Remarks on computing the probability integral in one and two dimensions", 1st Berkeley Symposium on Mathematical Statistics and Probability, 1945,63-78.

[18] B.C. M. Reddya, K. H. Reddyb , C. N. Muni Reddy, K.V. K. Reddy,"Quota Allocation to Distributors of the Supply Chain under Distributors' Uncertainty and Demand Uncertainty by Using Fuzzy Goal Programming", Jordan Journal of Mechanical and Industrial Engineering,Vol. 2, No. 4, 2008, 215 -226.

[19] K. M. Aljeborya, M. Alshebeb, "Integration of Statistical and Engineering Process Control for Quality Improvement (A Case Study: Chemical Industry - National Chlorine Industries)", Jordan Journal of Mechanical and Industrial Engineering,Vol. 8, No., 4, 2014, 243 – 256.

[20] P..S. Raoa, C..Ratnam, "Damage Identification of Welded Structures Using Time Series Models and Exponentially Weighted Moving Average Control Charts", Vol. 4, No.6, 2010, 701 – 710.

[21] M. Alata, W. Masarweh, S. Kamal, "A Fuzzy Monitoring System for an Extrusion Line",Jordan Journal of Mechanical and Industrial Engineering, Vol. 1, No 1, 2007, 17 – 21.

[22] F. Abdul Moneim, "Expected Delays in Completing Projects under Uncertainty", Jordan Journal of Mechanical and Industrial Engineering, Vol. 2, No 1, 2008, 65 – 69.

[23] M. Al-Marsumi, "Total Quality Management in the Top Rank of the Dairy Industry in Jordan", Jordan Journal of Mechanical and Industrial Engineering Mechanical and Industrial Engineering, Vol. 3, No 1, 2009, 47- 58.

[24] M. Al-Ghandoor, M. Samhouri, "Electricity Consumption in the Industrial Sector of Jordan: Application of Multivariate Linear Regression and Adaptive Neuro-Fuzzy Techniques", Jordan Journal of Mechanical and Industrial Engineering, Vol. 3, No 1, 2009, 69- 76.

[25] V. Raviprakash, A. P. Sidharth, B. Prabu, N. Alagumurthi,"Structural Reliability of Thin Plates with Random Geometrical Imperfections Subjected to Uniform Axial Compression", Jordan Journal of Mechanical and Industrial Engineering Vol. 4, No 2, 2010, 270- 279.

[26] Neelufur, K. S. Rao, K. V. Subbaiah, "Studies On $\bar{X}$- Control Chart With Pareto In-Control Times for Non Normal Variates",Jordan Journal of Mechanical and Industrial Engineering, Vol. 4, No 3, 2010, 364 – 371.

[27] P. S. Rao, C. Ratnam, "Damage Identification of Welded Structures Using Time Series Models and Exponentially Weighted Moving Average Control Charts",Jordan Journal of Mechanical and Industrial Engineering, Vol 4, No 6, 2010, 701 – 710.

[28] S. Prasad, M. M. Sarcar, "Analysis of Face Milling Operation Using Acousto Optic Emission and 3D Surface Topography of Machined Surfaces for In-Process Tool Condition Monitoring", Jordan Journal of Mechanical and Industrial Engineering, Vol 5, No 6, 2011, 509 – 519.

[29] O. Alsaed , I. S. Jalham, "Polyvinyl Butyral (PVB) and Ethyl Vinyl Acetate (EVA) as a Binding Material for Laminated Glass",Jordan Journal of Mechanical and Industrial Engineering, Vol 6, No 2, 2012, 127 – 133.

[30] Luo , S. Jing, Xi. Han, Y. Liu, C. Du, "Load Characteristics of Pick Cutting Coal Seams with Coal and Rock Interface",Jordan Journal of Mechanical and Industrial Engineering,Vol. 10 No 3,2016, 205-210.