ISSN 2347-1921

INFORMATION MEASURES AND SOME DISTRIBUTION APPROXIMATIONS

Mohammed E. M. Gobar, Eihab B.M. Bashier, Adnan M. Awad

Department of Mathematics, Faculty of Science and Arts, Baljurashi, Albaha University, Kingdom of Saudi Arabia

mudeer.all@gmail.com

Department of Mathematics, Faculty of Science and Arts, Baljurashi, Albaha University, Kingdom of Saudi Arabia

eihabbashier@gmail.com

Department of Mathematics, University of Jordan, 11942,

Jordan awada@gu.edu.jo

ABSTRACT

The Fisher and Kullback- Liebler information measures were calculated from the approximation of a binomial distribution by both the Poisson and the normal distributions and are applied to the approximation of a Poisson distribution by a normal distribution. In this paper the concept of relative loss in information due to approximating the distribution of a random variable X_n by that of another distribution of Y_n is introduced, and this concept is used to determine the value of the sample size for which the relative loss in information measure is less than a given level ϵ.

2000 Mathematics Subject Classification: 94A17

Keywords: Fisher information measures; Kullback- Liebler information measures; approximation distribution; relative loss.

Council for Innovative Research

Peer Review Research Publishing System

Journal: Journal of Advances in Mathematics

Vol 6, No. 3

editor@cirworld.com

www.cirworld.com, member.cirworld.com

ISSN 2347-1921

1. INTRODUCTION

Many information measures are suggested in the literature [1], [2],[4],[7],[9],[10] and [11]. Among these measures are the Fisher information measure (1925) [5] $I_F(\theta)$ that was introduced to measure the amount of information in a random variable X. This measure is defined by:

$$I_F(\theta) = -E\left(\frac{\partial^2\{\log g(x;\theta)\}}{\partial\theta^2}\right),$$

where, $g(x;\theta)$, $\theta \in \Omega$ is the probability density function $(p.d.f)$of the random variable. Kullback-Liebler(1956) [8] directed a divergence measure, which is defined as follows:

Let X be a r.v with $(p.d.f)$ $f_1(x)$ under the hypothesis H_1 or $f_2(x)$ under the hypothesis H_2. One of the main non-parametric measures of information is the Kullback-Liebler directed divergence measure, which is given by:

$$I(f_1, f_2) = E_{f_1}[\log\{f_1(X)/f_2(X)\}].$$

It is known that $Kl(f_1: f_2) \geq 0$ with equality holds iff $f_1(x) = f_2(x)$ almost surely. So, we select the sample size n such that $Kl(f_1: f_2)$ is minimum [6].

We will study the information embedded in a random variable X_n whose distribution for large n is approximated by the distribution of Y_n .

The question which arises here is that how large n should be for this approximation to be acceptable? To answer this question, we follow [3] and [6]. Let I be an information measure, we select the sample size n such that the relative loss in the Fisher and Kullback-Liebler information measures is less than some given small number ϵ ; $0 < \epsilon < 1$.

2. APPROXIMATIONS BY FISHER INFORMATION MEASURES

We measure information by using Fisher information measure :

2.1 Poisson Approximation to Binomial Distribution

Let $Y_1, \ldots, Y_n$ be i.i.d from $B(1,\theta)$, $X = \sum_{i=1}^{n} Y_i : B(n,\theta)$. The p.d.f of X is

$$f(x,\theta) = \begin{cases} \binom{n}{x}\theta^x(1-\theta)^{n-x} & if \quad x = 0,1,2,3,\ldots,n \\ 0 & otherwise \end{cases} \qquad (1)$$

Note that:

$$logf(x;\theta) = log(n!) - log(n-x)! - log(x!) + xlog\theta + (n-x)log(1-\theta)$$

From which,

$$\frac{\partial^2\{logf(x,\theta)\}}{\partial\theta^2} = -x/\theta^2 - (n-x)/(1-\theta)^2$$

Hence the Fisher information in X about θ is

$$I_{Bin}(\theta) = -E\left(\frac{\partial^2\{logf(x,\theta)\}}{\partial\theta^2}\right) = n/\theta + n/(1-\theta) \qquad (2)$$

Let X be approximated by $Y: P(n\theta)$ with a $p.d.f$

$$g(y;\theta) = \begin{cases} e^{-n\theta}(n\theta)^y/y! & if \quad y = 0,1,2,\ldots \\ 0 & otherwise \end{cases} \qquad (3)$$

Note that: $\log(g(y;\theta)) = -n\theta + ylog(n\theta) - log(y!)$

From which,

$$\frac{\partial^2\{\log g(y;\theta)\}}{\partial\theta^2} = -\frac{y}{\theta^2}$$

Hence the Fisher information in Y about θ is

$$I_P(\theta) = -E\left(\frac{\partial^2\{\log g(y;\theta)\}}{\partial\theta^2}\right) = n/\theta \qquad (4)$$

So, the loss in the Fisher information due to approximating X by Y is

$$I_{Bin}(\theta) - I_P(\theta) = n/(1-\theta).$$

It is strange that this loss increases with the sample size. Moreover, the relative loss in the Fisher information measure due to this approximation is

$$\boldsymbol{L} = |(I_{Bin}(\theta) - I_P(\theta))/I_{Bin}(\theta)| = |(n/\theta + n/(1-\theta) - n/\theta)/(n/\theta + n/(1-\theta))| = \theta$$

which does not depend on sample size n. Therefore, the relative loss in the Fisher information measure cannot be used to specify the sample size n for which the above approximation is suitable.

2.2 Normal Approximation to the Poisson Distribution

Let $Y_1, \dots, Y_n$ be i.i.d from $P(\theta)$ then $X = \sum_{i=1}^{n} Y_i : P(n\theta)$ with a $p.d.f$ as in Equation (3). The Fisher information of X is given by Equation (4), namely,

$$I_P(\theta) = n/\theta.$$

Let X be approximated by $Y: N(n\theta, n\theta)$ with a $p.d.f$

$$g(y,\theta) = \{1/\sqrt{2\pi n\theta}\} exp(-(y-n\theta)^2/2n\theta) \;\; , -\infty < y < \infty \;\; and \;\; 0 \; otherwise \qquad (5)$$

Note that: $log(g(y,\theta)) = -(1/2)log(2n\pi) - (1/2)log(\theta) - (y-n\theta)^2/2n\theta$

From which,

$$\frac{\partial^2 \log g(y,\theta)}{\partial \theta^2} = 1/2\theta^2 - y^2/n\theta^3$$

Hence the Fisher information in Y about θ is

$$I_N(\theta) = -1/2\theta^2 - E\,y^2/n\theta^3$$

But $EY^2 = var(Y) + (E(Y))^2 = n\theta + n^2\theta^2 = n\theta(1+n\theta)$

Hence, $I_N(\theta) = -1/2\theta^2 + n\theta(1+n\theta)/n\theta^3 = (1+n\theta)/\theta^2 - 1/2\theta^2 = (1+2n\theta)/2\theta^2$

Therefore, the relative loss in the Fisher information due to this approximation is

$$L = |(I_P(\theta) - I_N(\theta))/I_P(\theta)| = |\{n/\theta - [(1+2n\theta)/2\theta^2]\}/(n/\theta)| = 1/2n\theta$$

For a given small positive real number ε the relative loss $L < \varepsilon$ implies that

$$1/2n\theta < \varepsilon \qquad (6)$$

2.3 Normal Approximation to the Binomial Distribution

Let $X: B(n,\theta)$ with a $p.d.f$ as in Equation (1) .The Fisher information of X is given by $I_{Bin}(\theta) = n/\theta + n/(1-\theta)$ as given in Equation (2) above.

Let X be approximated by $Y: N(n\theta, n\theta(1-\theta))$ with a $p.d.f$

$$g(y,\theta) = \left\{1/\sqrt{2\pi n\theta(1-\theta)}\right\} \exp\{-(y-n\theta)^2/2n\theta(1-\theta)\} \qquad (7)$$

Note that $log(g(y,\theta)) = (-1/2)log(2\pi) - (1/2)\log n\theta(1-\theta) - \{(y-n\theta)^2/2n\theta(1-\theta)\}$

From which,

$$\frac{\partial^2 \log g(y,\theta)}{\partial \theta^2} = -[(-2\theta(1-\theta) - (1-2\theta)^2)/2\theta^2(1-\theta)^2] - \left[\begin{matrix} 2n\theta^2(1-\theta)^2(2y^2 + 2n^2\theta - 4n\theta y) - \\ (2y^2\theta - y^2 + n^2\theta^2 - 2n\theta^2 y) \\ (-4n\theta^2 + 4n\theta^3 + 4n\theta(1-\theta)^2) \end{matrix}\right] \Big/ 4n^2\theta^4(1-\theta)^4$$

Hence, the Fisher information in Y is

$$I_N(\theta) = [(-2\theta + 2\theta^2 - 1 + 4\theta - 4\theta^2)/2\theta^2(1-\theta)^2] - \left[\begin{matrix} 2n\theta^2(1-\theta)^2(2E(y^2) + 2n^2\theta - 4n\theta E(y)) \\ -(2E(y^2)\theta - E(y^2) + n^2\theta^2 - 2n\theta^2 E(y)) \\ (-4n\theta^2 + 4n\theta^3 + 4n\theta(1-\theta)^2) \end{matrix}\right] \Big/ 4n^2\theta^4(1-\theta)^4$$

Since $EY = n\theta$, $EY^2 = n\theta - n\theta^2 + n^2\theta^2$.

Hence,

$$I_N(\theta) = [(2\theta - 2\theta^2 - 1)/2\theta^2(1-\theta)^2] - \left[\begin{matrix} 2n\theta^2(1-\theta)^2(2n\theta - 2n\theta^2 + 2n^2\theta - 2n^2\theta^2) \\ -(3n\theta^2 - 2n\theta^3 - n\theta)(-12n\theta^2 + 8n\theta^3 + 4n\theta) \end{matrix}\right] \Big/ 4n^2\theta^4(1-\theta)^4$$

Then the relative loss in the Fisher information due to this approximation

$$L = |(I_{Bin}(\theta) - I_N(\theta))/I_{Bin}(\theta)| \qquad (\mathbf{8})$$

3. APPROXIMATIONS BY KULLBACK – LIEBLER INFORMATION MEASURES

We measure information by using Kullback- Liebler information measure :

3.1 Poisson Approximation to Binomial Distribution

Let $Y_1, \dots, Y_n$ be i.i.d from $B(1,\theta)$, $X = \sum_{i=1}^{n} Y_i : B(n,\theta)$ with a $p.d.f$ as in Equation (1) . On the other hand, assume that $X: P(n\theta)$ with a $p.d.f$ as in Equation (3); then Kullback-Liebler information measure is given by

$$Kl(f_1 : f_2) = E_{f_1} log\{f_1(x)/f_2(x)\} = E_{f_1} log\left[n!\, \theta^x ((1-\theta)^{n-x} x!)/x!\,(n-x)!\, e^{-n\theta}(n\theta)^x\right]$$

$$= log(n!) + n\theta log(\theta) + n(1-\theta) log(1-\theta) + n\theta - n\theta log(n\theta) - E_{f_1} log(n-x)! \qquad (9)$$

3.2 Normal Approximation to Poisson Distribution

Let $X = \sum_{i=1}^{n} Y_i : P(n\theta)$ with a $p.d.f$ as in equation (3). On the other hand let $X: N(n\theta, n\theta)$ with the $p.d.f$

$$f_0(x) = \{1/\sqrt{2\pi n\theta}\} \exp\{-(x-n\theta)^2/2n\theta\} \quad , -\infty < x < \infty \qquad (10)$$

It is interesting to note that $f_0(x)$ is a $p.d.f$ of a continuous distribution, the discrete analogue of this is

$$f_2(x) = P_{f_0}(x - 1/2 < X < x + 1/2) = \int_{x-1/2}^{x+1/2} f_0(y)dy = \{1/\sqrt{2\pi n\theta}\} \int_{x-1/2}^{x+1/2} \exp\{-(y-n\theta)^2/2n\theta\}dy \quad , \quad x = 0,1,2, \dots$$

Then the Kullback-Liebler information measure of this approximation is given by

$$Kl(f_1 : f_2) = E_{f_1} log\{f_1(x)/f_2(x)\} = \sum_{x=0}^{\infty} f_1(x) log\,(f_1(x)/f_2(x)) \qquad (11)$$

3.3 Normal Approximation to Binomial Distribution

Let $X = \sum_{i=1}^{n} Y_i : B(n,\theta)$ with a $p.d.f$ given by equation (1). On the other hand, let $X: N(n\theta, n\theta(1-\theta)$ with the $p.d.f$

$$f_0(x) = \left\{1/\sqrt{2\pi n\theta(1-\theta)}\right\} \exp\{-(x-n\theta)^2/2n\theta(1-\theta)\}\,, \quad -\infty < x < \infty \text{ and zero otherwise}$$

The discrete analogous p.d.f of this is

$$f_2(x) = P_{f_0}(x - 1/2 < X < x + 1/2)\,,\ x = 0,1,2, \dots$$

$$= \int_{x-1/2}^{x+1/2} f_0(y)dy = \left\{1/\sqrt{2\pi n\theta(1-\theta)}\right\} \int_{x-1/2}^{x+1/2} \exp\{-(y-n\theta)^2/2n\theta(1-\theta)\}dy$$

Then the Kullback-Liebler information measure is

$$Kl(f_1 : f_2) = E_{f_1} log\{f_1(x)/f_2(x)\} = \sum_{x=0}^{\infty} f_1(x) log\,(f_1(x)/f_2(x)) \qquad (12)\ .$$

4. RESULTS

Throughout this section, the Python 3.3 programs using the numpy library have been used to perform the computations and plotting the figures.

4.1 The Approximation by Fisher Information Measure

Equation (6) is used for computing the smallest sample size to approximate the Poisson distribution by the normal distribution, within a relative loss of $\varepsilon = 0.1$. The different values of the sample size corresponding to values of $0 < \theta < 1$ are explained in Figure 4.1 below.

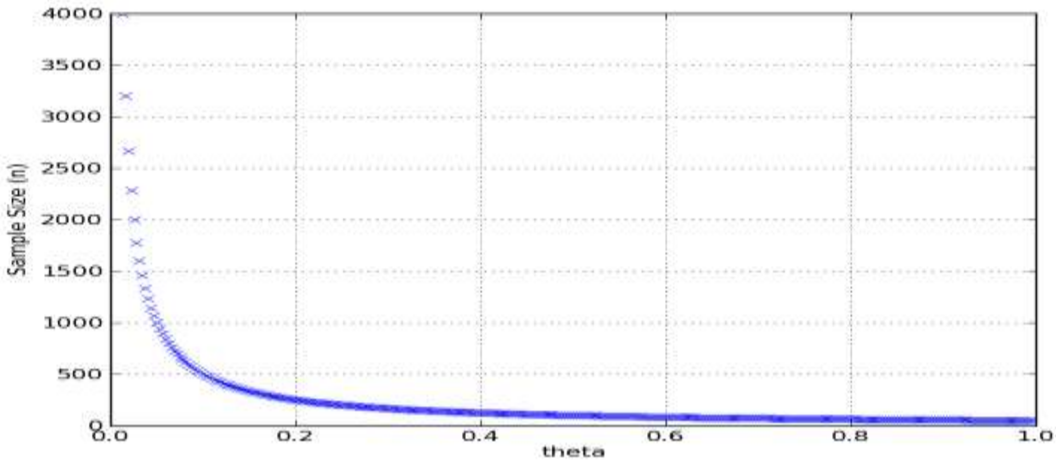

Figure 4.1: The smallest required sample size to approximate the Poisson distribution by the normal distribution for $0 < \theta < 1$, within a relative loss of $\varepsilon = 0.1$.

In Figure 4.2 below, we show the probability density functions for the Poisson and Normal distributions, for different values of the relative loss ($\varepsilon = 2^{-3}, 2^{-4}, \dots, 2^{-8}$) in the Fisher information measure.

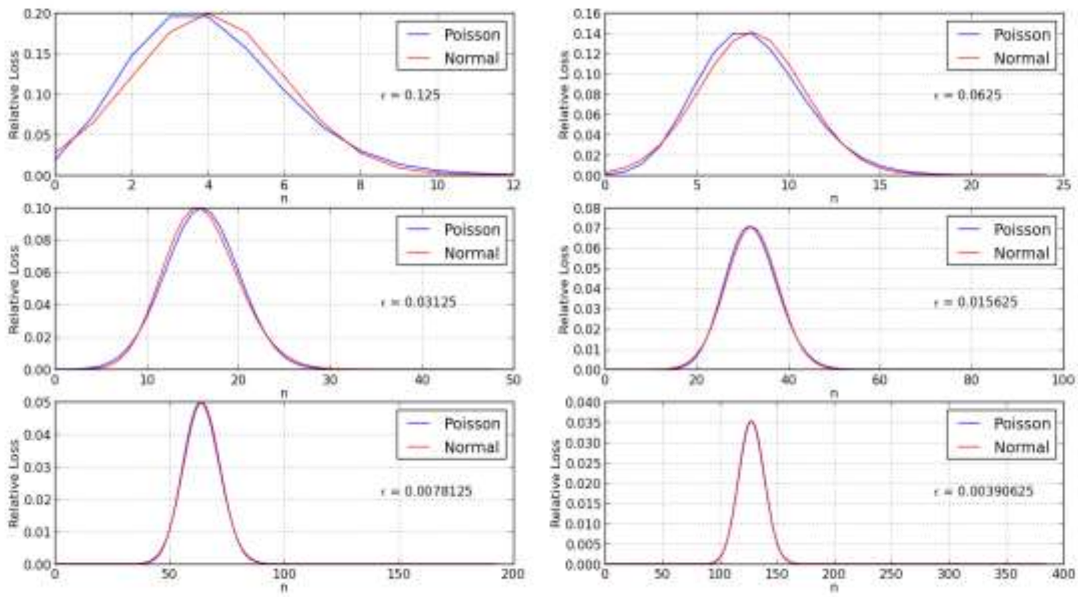

Figure 4.2: The probability density functions for the Poisson distribution and the normal distribution under the Fisher's approximation measure.

For the approximation of the binomial distribution by the normal distribution, Equation (8) is used to compute the value of the relative loss L as a function in the sample size n for each value of $\theta = 0.1{:}\,0.1{:}\,0.9$, for a relative loss $\varepsilon < 0.05$.

ISSN 2347-1921

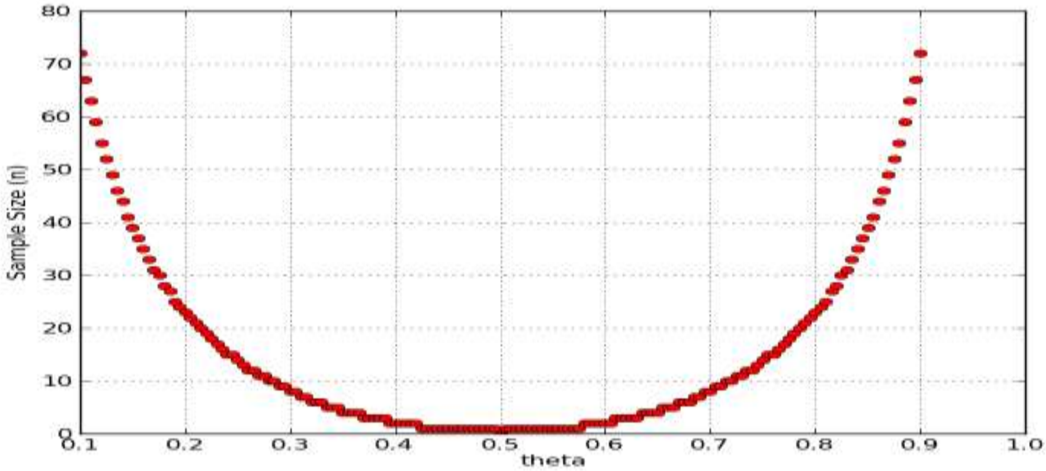

Figure 4.3: The required sample size to approximate the binomial distribution by the normal distribution for $\boldsymbol{\theta = 0.1:0.1:0.9}$, within a relative loss of $\boldsymbol{\varepsilon = 0.05}$

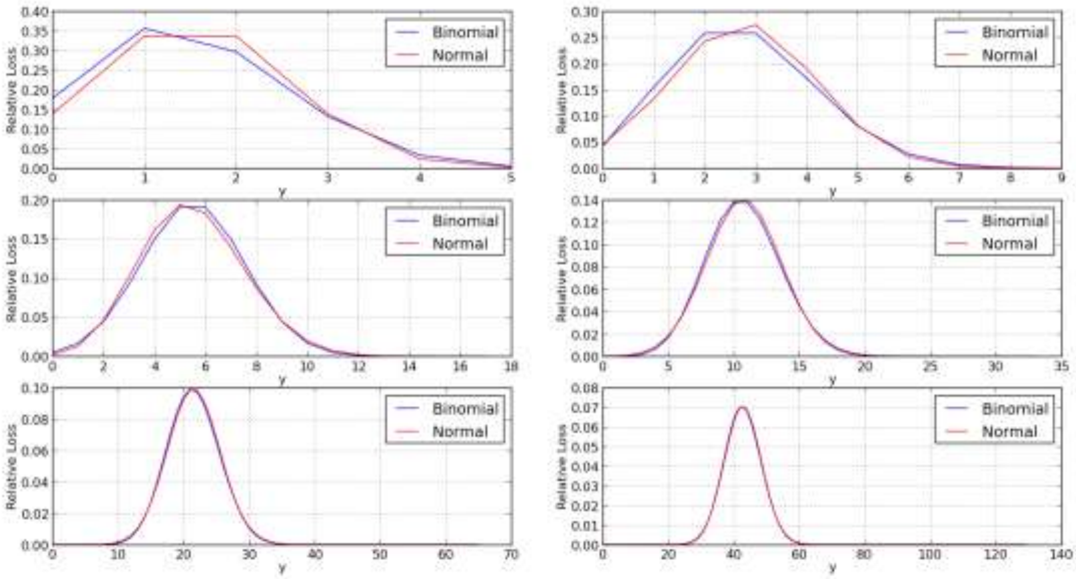

Figure 4.4: Approximation of the Binomial distribution by the normal distribution under the Fisher's information measure.

It is clear that Figure 4.3 give that this approximation is good enough even for small sample size i.e. $n \leq 20$ if $\theta = 0.2, 0.3, \ldots, 0.8$.

4.2 The Approximation by the Kullback-Liebler Information Measure

Equation (9) is used to evaluate $Kl(f_1 : f_2)$ as a function in the sample size n to approximate the Poisson distribution by the binomial distribution for $0.2 < \theta < 0.7$. The values of $Kl(f_1 : f_2)$ are explained in Figure 4.5.

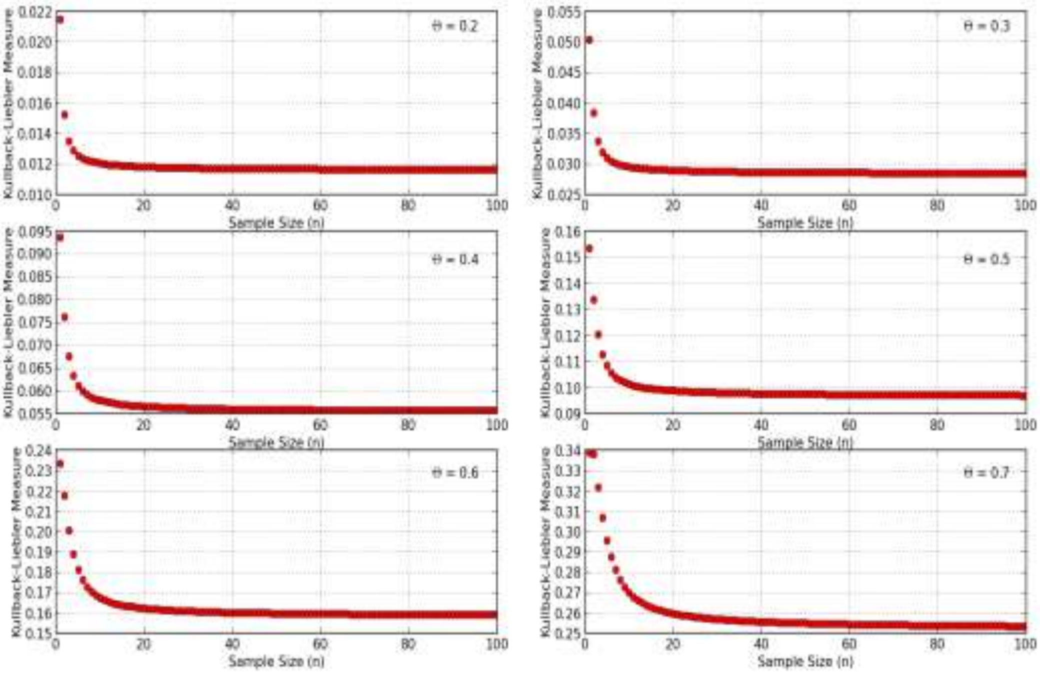

Figure 4.5: The Kullback-Liebler measure for approximating the Poisson distribution by the binomial distribution for $\theta = 0.2{:}\,0.1{:}\,0.7$

It is clear from the graph of Figure 4.5 that almost all values of $Kl(f_1{:}\,f_2)$ are very close to zero, which means that this approximation is very good even for small sample sizes.

To approximate the Poisson distribution by the normal distribution using the Kullback-Liebler information measure, Equation (11) is used to evaluate $Kl(f_1{:}\,f_2)$ for $\theta = 0.1{:}\,0.1{:}\,0.6$. The results obtained are explained in Figure 4.6 below.

ISSN 2347-1921

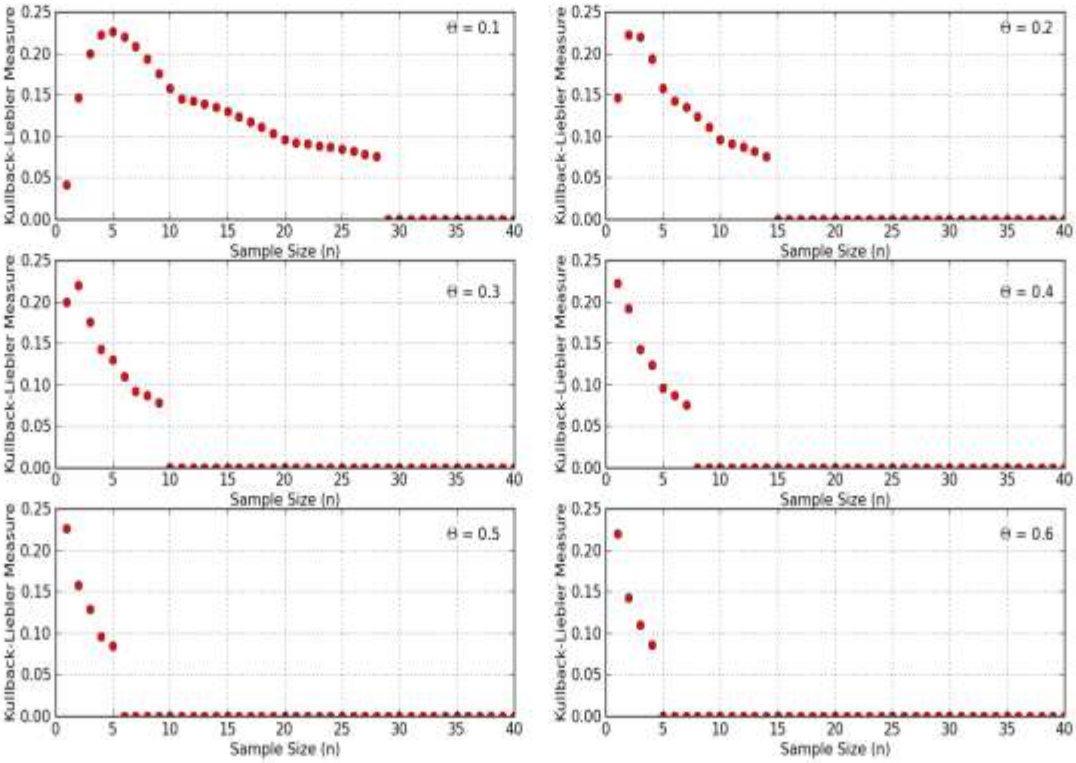

Figure 4.6: The Kullback-Liebler information measure for the approximation of the Poisson distribution by the normal distribution for $\theta = 0.1: 0.1: 0.6$

It is clear that from the Figure 4.6 that all values of $Kl(f_1: f_2)$ for the reported values of the sample size n are almost equal to zero . So the approximation is good for $n \geq 30$, and $\theta \geq 0.1$.

Finally, to approximate the binomial distribution by the normal distribution, Equation (12) is used to compute $Kl(f_1: f_2)$ for $\theta = 0.4: 0.1: 0.9$,Figure 4.7 shows the obtained values of $Kl(f_1: f_2)$ versus n.

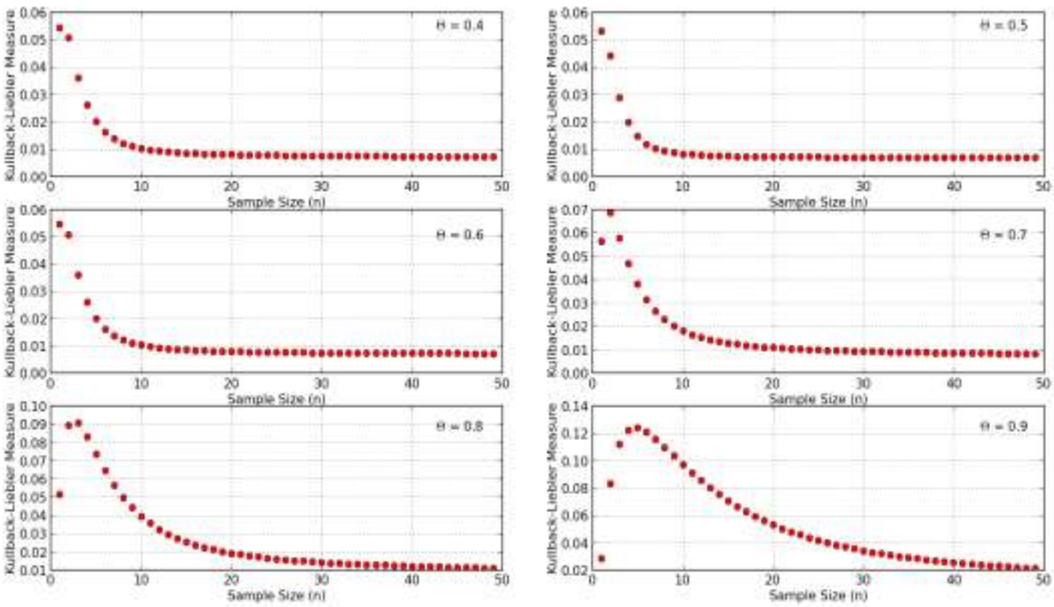

Figure 4.7: The Kullback-Liebler information measure for approximating the Binomial distribution by the normal distribution for $\theta = 0.4{:}\,0.1{:}\,0.9$

It is clear from Figure 4.7 that the smallest sample size for which the two distributions become within a tolerance value of $Kl(f_1{:}\,f_2) \leq 0.02$ increases with the value of the parameter θ. For $\theta = 0.4{:}\,0.1{:}\,0.9$ the approximation is good enough for values of $n \geq 50$.

5. CONCLUSIONS

This paper discussed the approximations of the Poisson distribution by both the binomial and normal distribution and the binomial distribution by the normal distribution using the Fisher and Kullback-Liebler information measures.

In the Fisher information measure, it was first noticed that this measure is not possible to approximate the Poisson distribution the normal distribution, since the relative loss is free of the sample size n. Then, from Figure 4.1 it was clear that the approximation Poisson distribution by the normal distribution is good enough even for small sample size when $0.2 < \theta < 0.8$. Note that for the cases reported in Figure 4.1 , the mean of the Poisson distribution $n\theta$ takes values between 20 and 80. These results are confirmed by Figure 4.2. Finally, when approximating the binomial distribution by the normal distribution, it is noticed from Figure 4.3 that the sample size is symmetric around $\theta = 0.5$. The sample size decreases in the range $0.1 < \theta < 0.5$ and increases in the range $0.5 < \theta < 0.9$. The binomial distribution and the normal distribution become identical for $\theta = 0.5$.

In the Kullback-Liebler information measure, when approximating the Poisson distribution by the binomial distribution, Figure 4.5 showed that the Kullback-Liebler measure increases proportional to θ. When approximating the Poisson distribution by the normal distribution, the smallest sample size for which the two approximation are identical decrease with the values of θ. The two distributions are always identical for sample sizes $n \geq 30$. Finally, when fixing a tolerance level $Kl(f_1{:}\,f_2) \leq 0.02$ it was noticed that the minimum sample size such that the Kullback-Liebler information measure of the approximation of the binomial distribution to the normal distribution become within the given tolerance increases with the value of θ, and if the sample size exceeds 50 the approximation is good enough.

REFERENCES

[1] A . M. Awad, and A.Abu-Taleb, Information relationship measures , IMA Journal of mathematical control and information ,4,1, 13-23,1987.

[2] A . M. Awad ,The Shannon entropy of generalized gamma and of related distributions : Proc. First Jordanian mathematics Conf. Yarmouk university. Jordan pp.13-27 ,1991.

[3] M.M.Azzam,and A . M. Awad, Entropy Measures And Some Distribution Approximations, *Microelectron, Reliab*. Vol.36(10):1569-1580, 1996.

[4] M.M.Azzam and A . M. Awad, Application Of The Shannon Entropy To Type II Censoring In Life Testing Experiments, *Journal of Fractional Calculus* Vol.11, May :103-112, 1997.

[5] R.A.Fisher , Proc.Cambridge Phil. Soc.22,700(1929) reprinted in collected papers of R.A.Fisher, edited by J.H.Bennett (Univ. of Adedlaide Press , South Australia, 1972),pp. 15-40.

[6] M.E.Gobar , Optimal sample size in approximating some sampling distribution , M . SC. ,Thesis , Jordan University 1999, Jordan.

[7] V.L.Hartley , Transmission of Information , *Bell System Tech.* J. Vol. 7 : 535-563 . 1928

[8] S.Kullback ,Information theory and statistics .Dover,U.S.A (1978) .

[9] A. Renyi, On Measures of Entropy and Information, *Proc. 4th Berkely Symp. Stat. and Pro*. 1960 ,Vol.1 *:547-561 .*

[10] C.A.Shannon, Mathematical Theory of Communication ,Bell System Tech. J., Vol.27(3):379-423, 1948

[11] A.D.Zighed ,G.Ritschard and S.Marcellin, Asymmetric and sample size sensitive entropy measures for supervised learning ,Advances in intelligent information systems . Vol. 265, 27-42, 2010.