American Journal of Scientific Research
ISSN 1450-223X Issue 33(2011), pp. 46-50

http://www.eurojournals.com/ajsr.htm

Optimal θ for Approximating the CDF of a Poisson Distribution by the Skew-Normal Distribution

R. M. Sakia
Department of Statistics, University of Botswana
Private Bag 00705, Gaborone, BOTSWANA
E-mail: sakiar@mopipi.ub.bw

Abstract

The Poisson density function and its CDF is widely taught in most introductory statistical courses. Before the advent of powerful computing power, evaluation of the CDF of the Poisson density for large values of k could only be obtained by the limiting large sample asymptotic normal approximation. In this paper we propose to approximate the Poisson CDF by using the Skewed-Normal distribution. Results show that this approximation gives far better performance than the asymptotic normal for both large and small k for 25≤θ≤50. It is slightly outdone by the Wilson-Hilferty approximation in this range. For values of θ>50 however, the three approximations seem to be very accurate in terms of both the absolute error and the minimal absolute error.

Keywords: Poisson CDF, Skewed-Normal, Wilson-Hilferty approximation, minimal absolute error

1. Introduction

The Poisson distribution is one of the simplest and widely applicable probability mass function encountered in statistics. It is also taught in basic statistical courses for undergraduate students in various fields of study. In agriculture for example, the spatial distribution of weeds in a rice field or of lesions per leaf may be suspected to have a Poisson distribution. For X~ Poisson (θ), θ>0 the corresponding probability mass function and distribution functions are respectively given as;

$$f_\theta(x)=\frac{e^{-\theta}\theta^x}{x!},\ x=0,1,........ \tag{1}$$

$$F_\theta(k)=P(X\le k)=\sum_{x=0}^{k}\frac{e^{-\theta}\theta^x}{x!},\ \text{for k=0, 1, …….} \tag{2}$$

One of the most cited approximations to the Poisson distribution is the large sample asymptotic normal approximation to the CDF (see for example, Rees (1987)) given as;

$$F_\theta(k)\simeq\phi\left(\frac{k+0.5-\theta}{\sqrt{\theta}}\right)\ \text{for } \theta\ge 10 \tag{3}$$

This approximation is taught in almost every basic course in statistics. Despite the current availability of powerful computer subroutines to evaluate the CDF of any probability density function, there is a need to highlight the usefulness of these approximations as a vital tool in the teaching and understanding the underlying relationships as well as shortcomings in their application.

An improvement to this approximation has been proposed by Wilson and Hilferty (1951) and referred to as the Wilson-Hilferty approximation in the form;

$$F_{\theta}(k) \simeq 1-\phi\left(\frac{c-\mu}{\sigma}\right) \tag{4}$$

where $c=\left[\frac{\theta}{1+k}\right]^{\frac{1}{3}}$ $\mu=1-\frac{1}{9(k+1)}$ and $\sigma=\frac{1}{3\sqrt{1+k}}$

A comparison of these two approximations was presented by Lesch and Jeske (2009) of which the Wilson-Hilferty was found to outperform the asymptotic normal approximation. In this paper, we seek to search for the optimum parameter for the Poisson distribution which gives the best approximation of the CDF for a range of values of k. We seek to do this by employing the distributional properties of the Skew-Normal distribution. Ideally, we shall look for a value of the parameter θ for which the absolute error for a given k is closest to zero.

2. The Skew-Normal Distribution

The distribution was introduced by O'Hagan and Leonard (1976) and has since been widely investigated by Azzalini (1985). Its characteristic properties have been studied by Arnold and Lin (2004) in detail. A random variable X is said to have the skew-normal distribution with location parameter μ, a scale parameter σ and a shape (skew) parameter λ denote as SN (μ, σ^2, λ) if the pdf of X is given as;

$$f(x:\mu,\sigma,\lambda)=\frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\left(\frac{x-\mu}{\sigma}\right)\right) \tag{5}$$

where ϕ and Φ are the standard normal pdf and cdf respectively.

Note that for λ=0 the resulting pdf is N(μ, σ^2). If λ<0 then the distribution is negatively skewed whereas for λ>0 positive skewness is recorded. The moment generating function of the distribution is given as;

$$M(t)=2\exp\left(\frac{t^2}{2}\right)\Phi\left(\frac{\lambda t}{\sqrt{1+\lambda^2}}\right) \tag{6}$$

From (6) the three central moments of the Skew-Normal distribution were given by Pewsley (2000) as follows;

$$E(X)=\mu+\lambda\sigma\sqrt{\frac{2\lambda^2}{\pi(1+\lambda^2)}} \tag{7}$$

$$E\left(X-E(X)\right)^2=\sigma^2\left(1-\frac{2\lambda^2}{\pi(1+\lambda^2)}\right) \tag{8}$$

$$E\left(X-E(X)\right)^3=\sigma^3\sqrt{\frac{2}{\pi}}\left(\frac{\lambda^2}{1+\lambda^2}\right)^{\frac{3}{2}}\left(\frac{4}{\pi}-1\right) \tag{9}$$

3. The Poisson Approximation by the Skew-Normal Distribution

The corresponding three central moments for a Poisson distribution are all equal to θ. Equating these to those of the skew-normal distribution and solving for μ, σ, and λ in terms of θ, the following results were obtained;

$$\mu = \theta - \left[\frac{2\theta}{4-\pi} \right]^{\frac{1}{3}} \quad (11)$$

$$\sigma = \sqrt{\theta + \left[\frac{2\theta}{4-\pi} \right]^{\frac{2}{3}}} \quad (12)$$

$$\lambda = \sqrt{\frac{\pi}{\left[2\theta(4-\pi)^2 \right]^{\frac{1}{3}} + (2-\pi)}} \quad (13)$$

The skew-normal approximation will then be given as;

$$\boldsymbol{F_\theta(k)} \simeq \int_{-\infty}^{k+0.5} \boldsymbol{f(x:\mu,\sigma,\lambda)dx} \quad (10)$$

For a given value of k and θ, the R Statistical Package will be used to evaluate the cdf $F_\theta(k)$ in (9). This will then be compared with the two approximations given in (3) and (4) above. Figures 1 a, b below gives the absolute errors for the asymptotic normal, Wilson-Hilferty and Skew-Normal approximations for k=5 and k=20 respectively. In both cases, the Wilson-Hilferty gives the least absolute errors followed closely by the skew-normal. For k=5 and θ≥25 and for k=20 and θ≥50 the absolute errors are zero.

Figure 1a: Plot of absolute errors for k=5 for the three approximation

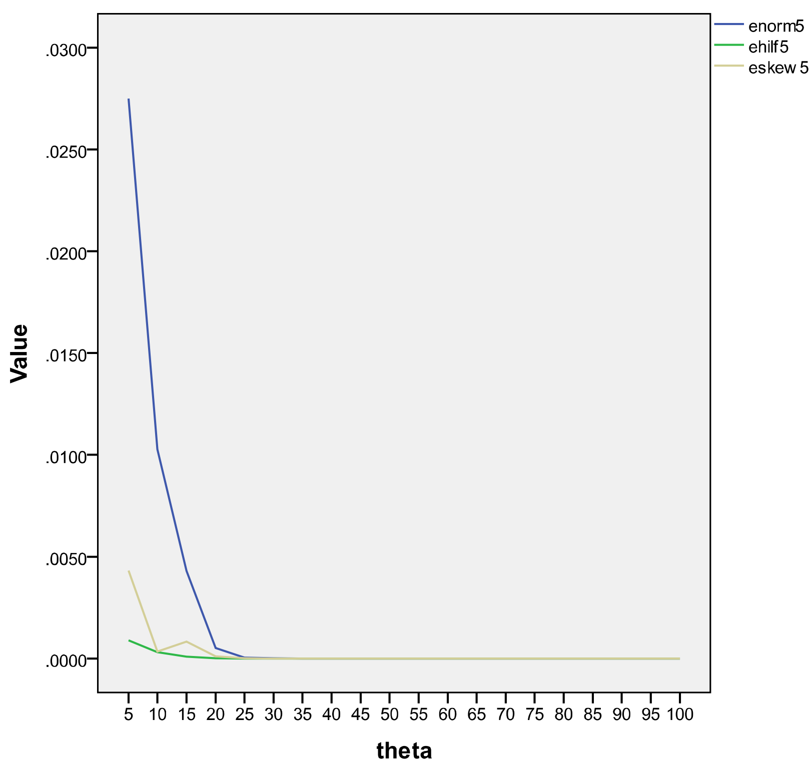
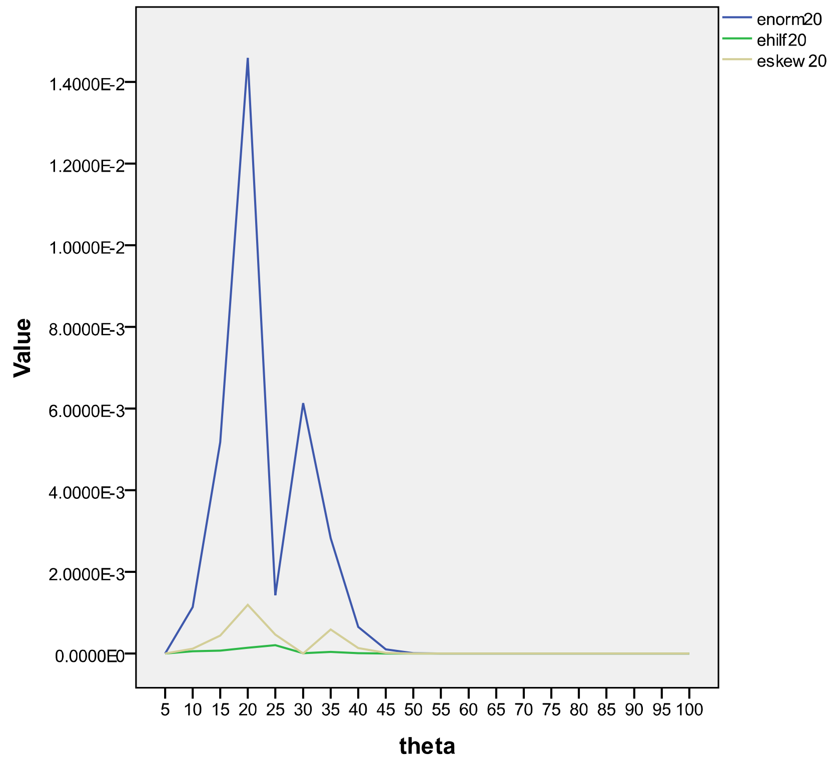
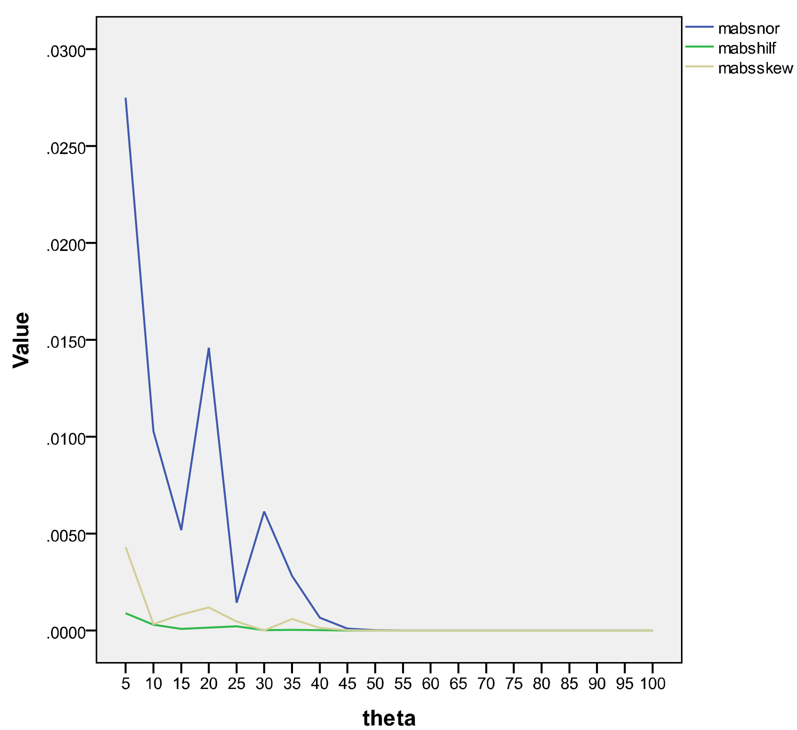


Figure 1b: Plot of absolute errors for k=20 for the three approximations



The maximal absolute errors of the approximation as defined by Schader and Schmid (1989) are presented in Fig. 2. Again, the Wilson –Hilferty approximation gives the best performance and closely followed by the Skew-Normal approximation.

Figure 2: Maximal absolute errors for the three approximations



The results for the Skew-Normal approximation indicates that for small k a parameter $\theta \geq 25$ gives a very accurate estimate of the CDF of the Poisson distribution. For a large k however, similar results are achieved for a parameter value $\theta \geq 45$. In general, the results for the Skew-Normal and Wilson-Hilferty for the approximation of the Poisson CDF are very close. In terms of the absolute maximal errors, all three approximations give very accurate approximations for the Poisson CDF for a parameter $\theta \geq 50$.

References

[1] Arnold, B.C. and Lin, G.D. (2004). Characterizations of the Skew-Normal and generalized Chi distributions. Sankya: The Indian Journal of Statistics, 66, 593-606.

[2] Azzalini, A. (1985). A class of distributions which includes the Normal ones. Scandinavian Journal of Statistics, 12, 171-178.

[3] O'Hagan, A. and Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. Biometrika, 63, 201-203

[4] Rees, D.G.(1987). Foundations of Statistics. Chapman and Hall

[5] Schader , M. and Schmidt, F. (1989)Two rules of thumb for the approximation of the Binomial distribution by the Normal distribution. The American Statistician, 43, 23-24

[6] Wilson, E.B. and Hilferty, M.M. (1931). The distributions of Chi-square. Proceedings of the National Academy of Sciences of the USA, 17,684-688.