A Note on Improved Approximation of the Binomial Distribution by the Skew-Normal Distribution

Ching-Hui CHANG, Jyh-Jiuan LIN, Nabendu PAL, and Miao-Chen CHIANG

It is a common practice to approximate a binomial distribution by a suitable normal distribution when n, the number of trials, is moderately large. But when p, the probability of success, is not close to 0.5, the binomial distribution can be heavily skewed, and hence the usual normal approximation may not be a good idea. In this note we show that the skew-normal distribution can provide a far better approximation due to its flexibility, and it can be used to approximate distributions other than the binomial one.

KEY WORDS: Central limit theorem; Cumulative distribution function (cdf); Skew parameter.

1. INTRODUCTION

Suppose the random variable Y follows a binomial distribution with parameters n (= the number of trials) and p (= the probability of success); that is, $Y \sim B(n, p)$, $p \in (0, 1)$. Define $F_{n,p}(k)$ as $F_{n,p}(k) = P(Y \leq k) =$ cdf of $B(n, p)$ at k, $k \in \{0, 1, \ldots, n\} = R_Y$(say). For small n, one can compute $F_{n,p}(k)$ fairly easily using any calculator. But for moderate-to-large values of n, $F_{n,p}(k)$ is conveniently approximated by the normal cdf as

$$F_{n,p}(k) \approx \Phi((k + 0.5 - np)/\sqrt{np(1 - p)}), \qquad (1)$$

where $\Phi(\cdot)$ represents the standard normal cdf. The above approximation is widely covered in most of the statistics test books, especially the introductory level ones. The value 0.5 used on the right hand side of (1) is called the "boundary value adjustment" (or the "continuity correction"), since a smooth normal cdf is used to approximate a step function cdf of the binomial distribution.

Most of the introductory level books set conditions for using the approximation (1). The two most popular such conditions are: (i) $np(1 - p) > 9$; and (ii) $np > 5$ for $0 < p \leq 0.5$, $n(1 - p) > 5$ for $0.5 < p < 1$.

Ching-Hui Chang is Associate Professor of Statistics, Department of Applied Statistics & Information Science, Ming Chuan University, Taoyuan County, Taiwan, ROC. Jyh-Jiuan Lin is Associate Professor of Statistics, Department of Statistics, Tamkang University, Tamsui, Taiwan, ROC. Nabendu Pal is Professor of Statistics, Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA, USA 70504-1010. Miao-Chen Chiang is Research Assistant in the Department of Statistics, Tamkang University, Tamsui, Taiwan, ROC. E-mail address for correspondence: *chchang@mcu.edu.tw*. We would like to thank the editor, Professor Peter Westfall, and an anonymous associate editor for their comments and suggestions which greatly improved the presentation of this note.

Schader and Schmid (1989) investigated the "maximal absolute error (MABS)" of the approximation, defined as

$$\text{MABS}(n, p) = \max_{k \in R_Y} |F_{n,p}(k) - \Phi((k + 0.5 - np)/\sqrt{np(1 - p)})| \qquad (2)$$

under the above conditions which could be substantial when p is near 0 or 1.

By the central limit theorem (CLT), (Y/n) approximately follows a $N(p, p(1 - p)/n)$ distribution; that is, Y approximately follows $N(np, np(1-p))$, for large n. Though the CLT justifies using a normal distribution to approximate a binomial distribution, there may be a question of aptness, especially where n is not "large," because except for $p = 0.5$, the $B(n, p)$ distribution is a skewed one. Hence, the natural question is: "Is there a better way to approximate a binomial distribution?"

The skew-normal distribution, which is a nice generalization of the normal distribution, seems to be a natural choice to approximate the binomial distribution. The skew-normal distribution is more versatile than the normal distribution, and can take negatively skewed, symmetric, or positively skewed shape (just like a binomial distribution) through its skew parameter. The purpose of this note is to show that the skew-normal distribution provides a far better approximation of the binomial distribution than the usual normal one.

Section 2 provides a brief background of the skew-normal distribution and some important references. Section 3 shows why the skew-normal distribution should be used over the normal distribution for approximating the $B(n, p)$ distribution. We also mention in Section 4 how this can be extended for other distributions as well.

2. THE SKEW-NORMAL DISTRIBUTION: A BRIEF BACKGROUND

A random variable X follows a skew-normal distribution with location parameter μ, scale parameter σ, and skew parameter λ (henceforth called the SN(μ, σ^2, λ) distribution) if the pdf of X is given by

$$f(x|\mu, \sigma, \lambda) = (2/\sigma)\phi((x - \mu)/\sigma)\Phi(\lambda(x - \mu)/\sigma), \qquad (3)$$

where ϕ and Φ denote the standard normal pdf and cdf, respectively. The parameter space is: $\mu \in (-\infty,\infty)$, $\sigma > 0$, $\lambda \in (-\infty, \infty)$. The distribution SN$(0, 1, \lambda)$ is called the standard skew-normal distribution with pdf $f(x|0, 1, \lambda) = 2\phi(x)\Phi(\lambda x)$. When $\lambda = 0$, the above pdf (3) boils down to the $N(\mu, \sigma^2)$ pdf. Also $\lambda > 0$ (< 0) implies positively (negatively) skewed shape of $f(x|\mu, \sigma, \lambda)$ given above. The skew-normal distribution was first introduced by O'Hagan and Leonard (1976) as

 DOI: 10.1198/000313008X305359

a prior distribution for estimating a normal location parameter. Some basic properties of the standard skew-normal distribution (i.e. ,SN(0, 1, λ)) are:

1. if $Z \sim \mathrm{SN}(0,\ 1,\ \lambda)$, then $(-Z) \sim \mathrm{SN}(0,\ 1,\ -\lambda)$;
2. if $Z \sim \mathrm{SN}(0,\ 1,\ \lambda)$, then $Z^2 \sim \chi_1^2$ (chi-square with 1 df);
3. as $\lambda \to \pm\infty$, SN(0, 1, λ) tends to the half-normal distribution, that is, the $\pm|N(0,\ 1)|$ distribution; and
4. the moment generating function of SN(0, 1, λ) is $M(t|\lambda) = 2\Phi(\delta t)\exp(t^2/2)$, where $\delta = \lambda/\sqrt{1+\lambda^2}$ and $t \in (-\infty,\ \infty)$.

For more properties of the skew-normal distribution one can see Azzalini (1985); Gupta, Nguyen and Sanqui (2004); or Arnold and Lin (2004).

3. APPROXIMATION BY THE SKEW-NORMAL DISTRIBUTION

In this section we will see how the SN(μ, σ^2, λ) distribution can be used to approximate a $B(n, p)$ distribution. It is hoped that the extra (skew) parameter λ would provide a better approximation of the binomial distribution by the SN(μ, σ^2, λ) distribution for $p \neq 0.5$.

For a given $B(n,\ p)$ distribution, the parameters of the approximating (or, matching) SN($\mu,\ \sigma^2,\ \lambda$) distribution are found by the methods of moments, that is,

1. $E(Y) = np = E(X) = \mu + \sigma(\sqrt{2/\pi})(\lambda/\sqrt{1+\lambda^2})$;
2. $E(Y-E(Y))^2 = np(1-p) = E(X-E(X))^2 = \sigma^2\{1-(2/\pi)\lambda^2/(1+\lambda^2)\}$;
3. $E(Y-E(Y))^3 = np(p-1)(2p-1) = E(X-E(X))^3 = \sigma^3(\sqrt{2/\pi})(\lambda/\sqrt{1+\lambda^2})^3((4/\pi)-1)$.

The above moment expressions of the skew-normal distribution can be found in Pewsey (2000).

Solving the above three equations is straight forward. We obtain the values of λ, σ, and μ as follows.

(a) Given $(n,\ p)$, the above last two equations yield

$$\frac{\{1-(2/\pi)\lambda^2/(1+\lambda^2)\}^3}{(2/\pi)(\lambda^2/(1+\lambda^2))^3((4/\pi)-1)^2} = np(1-p)/(1-2p)^2. \tag{4}$$

Note that the above Equation (4) involves λ^2. Hence solve for λ^2 (> 0) using (4). Then λ is found as

$$\lambda = \{\text{sign of } (1-2p)\}\sqrt{\lambda^2}. \tag{5}$$

(b) Given (n, p) and after obtaining λ as above, σ is found as

$$\sigma = \sqrt{np(1-p)}\,/\{1-(2/\pi)\lambda^2/(1+\lambda^2)\}^{1/2}. \tag{6}$$

(c) Finally, after obtaining λ and σ as above, get μ as

$$\mu = np - \sigma(\sqrt{2/\pi})\lambda/\sqrt{1+\lambda^2}. \tag{7}$$

Remark 1. For $p = 0.5$, the above (c) forces λ to take the value 0, and so we get back to the usual normal approximation.

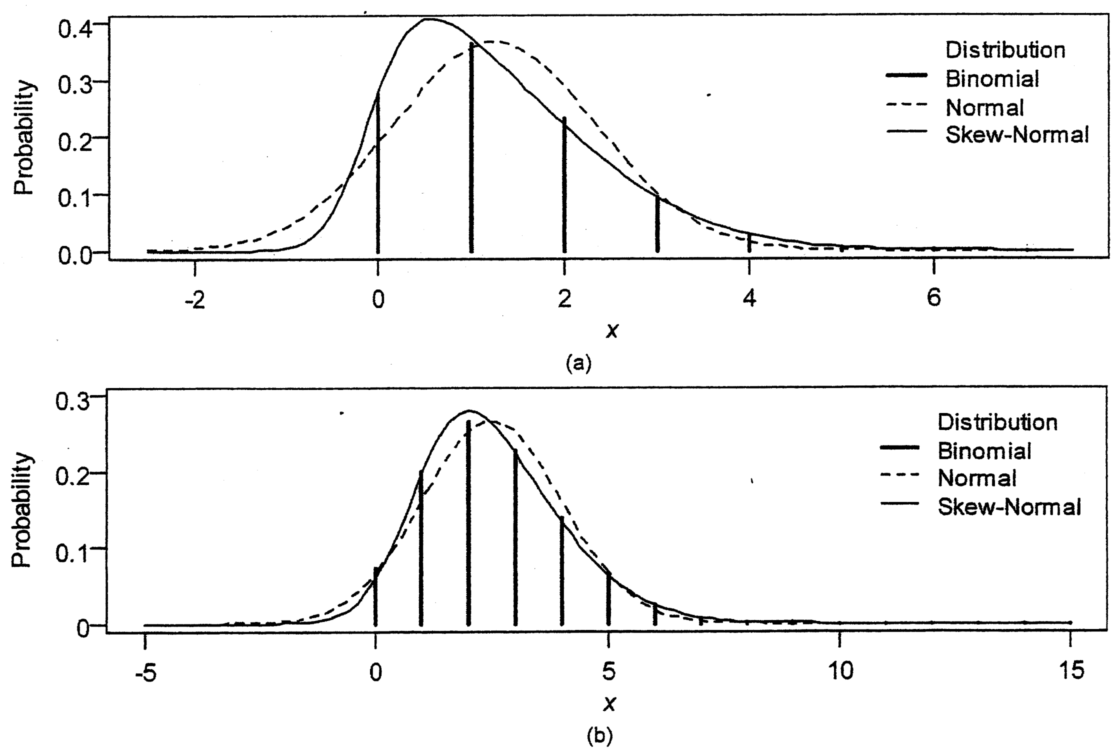

Figure 1. (a) $B(25,\ 0.05)$ pmf with the matching normal pdf and skew-normal pdf. (b) $B(25,\ 0.10)$ pmf with the matching normal pdf and skew-normal pdf.

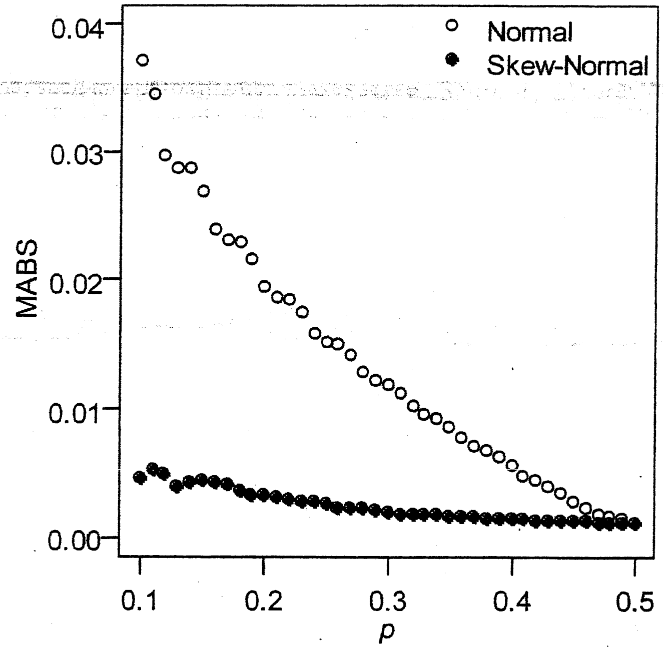

Figure 2. Plots of MABS(SN) and MABS(N) as functions of p ($n = 25$).

After obtaining λ, σ, and μ as given above, the binomial cdf is approximated as

$$F_{n,p}(k) \approx \int_{-\infty}^{k+0.5} f(x|\mu,\ \sigma,\ \lambda)dx = \Psi_\lambda((k+0.5-\mu)/\sigma), \tag{8}$$

where

$$\Psi_\lambda(c) = \int_{-\infty}^{c} 2\phi(z)\Phi(\lambda z)dz \tag{9}$$

is the standard skew-normal cdf.

The values of $\Psi_\lambda(c)$ can be obtained by Mathematica package, or an extended table for $\Psi_\lambda(c)$, $\lambda > 0$ can be made available by the corresponding author.

As a demonstration of the improved approximation of the binomial distribution by the skew-normal distribution, we first show the matching skew-normal and normal pdf's for the binomial distributions in Figure 1: Figure 1(a) for $(n, p) = (25, 0.05)$, and Figure 1(b) for $(n, p) = (25, 0.10)$. It can be seen that the skew-normal curve follows the asymmetric binomial distribution closely.

Next, we calculate the MABS (see (2)) over the interval $p \in (0, 0.5)$ for both the skew-normal as well as the normal approximations (henceforth called MABS(SN) and MABS(N), respectively). For fixed n, the behavior of MABS as a function of p is symmetric about $p = 0.5$. Figure 2 shows the plots of MABS(SN) and MABS(N) for $n = 25$. Plots for other values of n also showed identical patterns. The plots clearly show that skew-normal approximation is much superior to the usual normal approximation, especially when p is extreme. As expected, the two approximations get very close as p approaches 0.5.

Figure 3 shows the MABS plots for fixed p, but varying n. The trends are similar for other values of p also, though the distance between the curves get smaller. As expected, MABS diminishes with increasing n.

Remark 2. Though our proposed improved approximation of the binomial distribution works well for almost all n and p, it should be kept in mind that such an approximation makes sense for large n only. Also, there is a mild restriction on n and p to make the skew-normal approximation work. This is explained in the following paragraph.

Define $u = \lambda^2/(1+\lambda^2)$ and $v = 1/u \geq 1$. Then the left-hand side of (4) $= (1/k_*)(v - (2/\pi))^3$, where $k_* = (2/\pi)((4/\pi) - 1)^2$. Note that the left-hand side of (4) is increasing in $v \geq 1$. So, $\min_{v\geq 1}$ {left-hand side of (4)} $=$ {left-hand side of (4)}$|_{v=1}$ $= \{1-(2/\pi)\}^3/\{(2/\pi)((4/\pi)-1)^2\} \approx 1$ (actually 1.0095). But the right-hand side of (4) $= np(1-p)/(1-2p)^2 \to 0$ as $p \to 0$; $\to 0$ as $p \to 1$; $\to \infty$ as $p \to 0.5$. In order to get a solution of (4), we must have a restriction of (n, p) which is: right-hand side of (4) $\geq$ min of left-hand side of (4) ≈ 1; that is, the condition on $(n,\ p)$ is: $np(1-p)/(1-2p)^2 \geq 1$; that is, (after simplifications) $p^2(1+(4/n)) - p(1+(4/n)) + (1/n) \leq 0$. Fix n, then p must satisfy the last inequality in order to get a solution of (4). This inequality says that p must be between the roots which are $(1/2) \mp (1/2)(1+(4/n))^{-1/2}$, which is almost $(0, 1)$ for sufficiently large n. For example, with $n = 100$, the feasible interval for p becomes (0.0097, 0.9903) which enables us to apply the skew-normal approximation. For p very close to 0 or 1, neither the skew-normal will work, nor the normal approximation will be satisfactory. Hence one has to look for other approaches, perhaps Poisson approximation, which can make sense.

Remark 3. This section clearly shows how the skew-normal distribution can provide a much better approximation to binomial probabilities than the usual normal approximation. Also, the method presented here can easily be incorporated in statis-

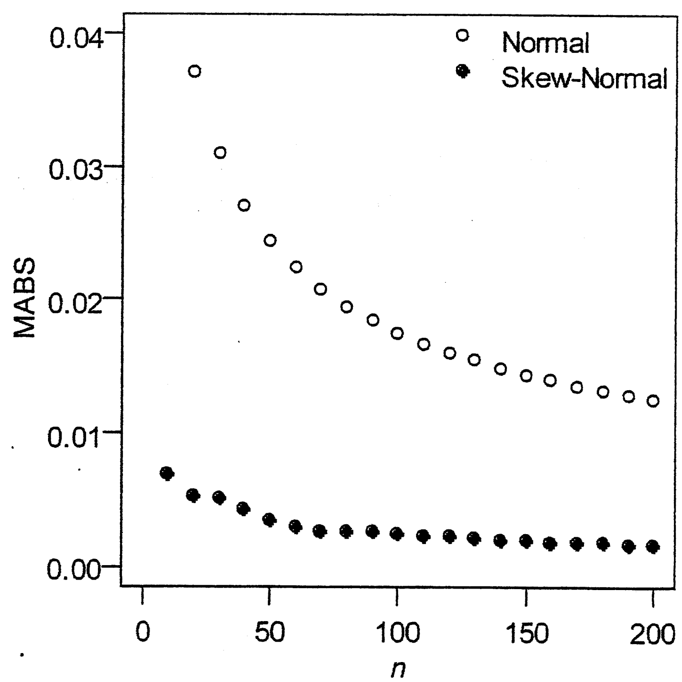

Figure 3. Plots of MABS(SN) and MABS(N) as functions of n ($p = 0.10$).

tical software to make this improved approximation applicable and user-friendly.

4. APPROXIMATIONS FOR OTHER DISTRIBUTIONS

The technique of skew-normal approximation as discussed in the previous section can be applied to other suitable distributions. For example, if the random variable Y follows NB(r, p) (negative binomial distribution where Y represents the number of failures before the rth success), then by the CLT one can approximate $P(Y \leq k)$ by $\Phi((p(k+0.5) - rq)/\sqrt{rq})$, $q = 1 - p$. Yet, a better approximating SN(μ, σ^2, λ) can be found by equating the first three moments of these two distributions as follows.

(a) Given $(r,\ p)$, find λ satisfying

$$\lambda = \{(\sqrt{rq}\sqrt{(2/\pi)}((4/\pi)-1)/(1+q))^{2/3} + (2/\pi) - 1\}^{-1/2}.$$

(b) Given (r, p) and after obtaining λ as above, σ is found as

$$\sigma = (\sqrt{rq}/p)/\{1 - (2/\pi)\,\lambda^2/(1+\lambda^2)\}^{1/2}\ .$$

(c) After obtaining λ and σ as above, get μ as

$$\mu = (rq/p) - \sigma(\sqrt{2/\pi})\,\lambda/\sqrt{1+\lambda^2}.$$

Our numerical computations show that the level of accuracy in the approximation provided by the skew-normal distribution in the negative binomial case is even greater than that of the normal distribution, way better than what we have seen in the binomial case. Also, similar to the binomial case, a mild restriction is put on r and p, which is $(1+q)/\sqrt{q} < \sqrt{r}/(1.00475)$, in order to make the skew-normal approximation work, that is, $P(Y \leq k) \approx \Psi_\lambda((k + 0.5 - \mu)/\sigma)$.

Similar skew-normal approximation can also be applied to a hypergeometric distribution, and details are omitted here.

[Received December 2007. Revised February 2008.]

REFERENCES

Arnold, B. C., and Lin, G. D. (2004), "Characterizations of the Skew-Normal and Generalized Chi Distributions," *Sankhyā: The Indian Journal of Statistics*, 66, 4, 593–606.

Azzalini, A. (1985), "A Class of Distributions Which Includes the Normal Ones," *Scandinavian Journal of Statistics*, 12, 171–178.

Gupta, A. K., Nguyen, T. T., and Sanqui, J. A. T. (2004), "Characterization of the Skew-Normal Distribution," *Annals of the Institute of Statistical Mathematics*, 56, 351–360.

O'Hagan, A., and Leonard, T. (1976), "Bayes Estimation Subject to Uncertainty about Parameter Constraints," *Biometrika*, 63, 201–203.

Pewsey, A. (2000), "Problems of Inference for Azzalini's Skew-Normal Distribution," *Journal of Applied Statistics*, 27, 859–870.

Schader, M., and Schmid, F. (1989), "Two Rules of Thumb for the Approximation of the Binomial Distribution by the Normal Distribution," *The American Statistician*, 43, 23–24.