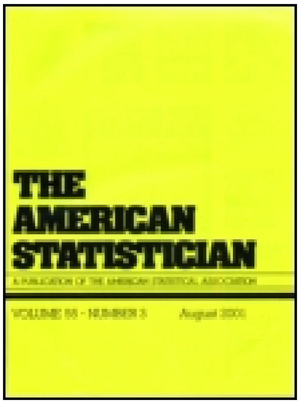

The American Statistician
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utas20

Some Suggestions for Teaching About Normal Approximations to Poisson and Binomial Distribution Functions

Scott M. Lesch and Daniel R. Jeske
Scott M. Lesch is Principal Consulting Statistician, Statistical Consulting Collaboratory, University of California, Riverside, CA 92521 . Daniel R. Jeske is Professor of Statistics and Chair, Department of Statistics, University of California, Riverside, CA 92521 .
Version of record first published: 01 Jan 2012.

To cite this article: Scott M. Lesch and Daniel R. Jeske (2009): Some Suggestions for Teaching About Normal Approximations to Poisson and Binomial Distribution Functions, The American Statistician, 63:3, 274-277

To link to this article: http://dx.doi.org/10.1198/tast.2009.08147

Some Suggestions for Teaching About Normal Approximations to Poisson and Binomial Distribution Functions

Scott M. LESCH and Daniel R. JESKE

In this article we review two historical approximations to the Poisson and binomial cumulative distribution functions (CDFs); that is, the Wilson–Hilferty and Camp–Paulson approximations. Both of these approximations reduce to standard normal formulas that produce very accurate estimates of the Poisson and binomial CDFs, and are thus quite simple to implement. Additionally, in an upper-division undergraduate or master's level probability and inference course, the derivation of these approximations presents a nice opportunity to introduce and study the distributional relationships between the gamma and Poisson CDFs, and the binomial, beta, and F CDFs. This article presents the basic theorems and lemmas needed to derive each approximation, along with some relevant examples that compare and contrast the precision of these approximations with their large-sample, limiting normal counterparts.

KEY WORDS: Camp–Paulson approximation; Distributional relationships; Normal approximation; Wilson–Hilferty approximation.

1. INTRODUCTION

The Poisson and binomial distributions are two of the simplest yet widely applicable probability mass functions used in statistics, and commonly discussed in many standard statistical textbooks. For $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$, the corresponding probability mass function ($f(x)$) and distribution function ($F(k)$) are defined to be $f_\lambda(x) = e^{-\lambda}\lambda^x/x!$, $x = 0, 1, 2, \ldots,$ and $F_\lambda(k) = \Pr(X \le k) = \sum_{x=0}^{k}(e^{-\lambda}\lambda^x/x!)$, for $k = \{0, 1, 2, \ldots\} = \Omega_X$. Likewise, for $Y \sim \text{Bin}(n, p)$, $y = 0, 1, \ldots, n$; $n \ge 1$, $p \in (0, 1)$, recall that $f_{n,p}(y) = \binom{n}{y} p^y (1-p)^{n-y}$, and $F_{n,p}(k) = \Pr(Y \le k) = \sum_{y=0}^{k} \binom{n}{y} p^y (1-p)^{n-y}$ for $k = \{0, 1, 2, \ldots, n\} = \Omega_Y$. Along with multiple examples of their use, many textbooks also present the well-known, large-sample asymptotic normal approximations to each cumulative distribution function (CDF); that is,

$$F_\lambda(k) \approx \Phi\big((k + 0.5 - \lambda)/\sqrt{\lambda}\big) \tag{1}$$

and

$$F_{n,p}(k) \approx \Phi\big((k + 0.5 - np)/\sqrt{np(1-p)}\big). \tag{2}$$

Scott M. Lesch is Principal Consulting Statistician, Statistical Consulting Collaboratory, University of California, Riverside, CA 92521 (E-mail: scott.lesch@ucr.edu). Daniel R. Jeske is Professor of Statistics and Chair, Department of Statistics, University of California, Riverside, CA 92521 (E-mail: daniel.jeske@ucr.edu).

Note that in (1) and (2), $\Phi(\cdot)$ represents the standard normal CDF and the "continuity correction" factor of 0.5 is used to improve each approximation.

In a modern statistical course where students are expected to have access to a computer, approximation formulas for either of these CDFs are in some sense unnecessary. Exact calculations for fairly large sample sizes can be easily obtained using most statistical software packages (or via direct programming, if necessary). Nonetheless, approximations can still be useful and relevant, both for simplifying large sample calculations and teaching various statistical principles such as the Central Limit Theorem (CLT) and/or various distributional relationships (i.e., the use of the Poisson CDF to approximate the binomial CDF, etc.). Approximations to these distributions can also be used to teach various moment-matching techniques; for example, see Chang et al. (2008).

Although (1) and (2) can certainly be viewed as useful for teaching CLT principles, neither approximation is particularly good. In this article, we review and discuss two other normal-based approximations to the Poisson and binomial CDFs. Both of these approximations possess three desirable properties. First, like (1) and (2), they each reduce to standard normal formulas and are thus simple to implement. Second, both approximations are known to be quite accurate and they each substantially outperform their classic large-sample counterparts. Third, they each can be (and indeed originally were) derived by exploiting some interesting distributional relationships. These various distributional relationships (between the gamma and Poisson CDFs, and the binomial, beta, and F CDFs) are often taught in advanced undergraduate and/or master's level probability and inference courses. Hence, the derivation of these approximations presents a nice instructional opportunity for introducing and/or reinforcing concepts about these types of distributional relationships.

In Section 2 of this manuscript we present the basic theorems and lemmas needed to derive each approximation. We also compare and contrast the precision of these approximations with their large-sample counterparts, and discuss a few relevant examples. Section 3 offers some brief instructional suggestions and concluding remarks.

It should be emphasized that our discussion about approximations for the Poisson and binomial CDFs is very narrowly focused and by no means comprehensive. Johnson and Kotz (1969) present a detailed, historical review of multiple approximation techniques for both of these CDFs. Likewise, Peizer and Pratt (1968) specifically discuss various types of normal-based approximations to the Poisson and binomial (as well as many other) distributions.

 DOI: 10.1198/tast.2009.08147

2. DERIVATION OF THE APPROXIMATIONS

The derivations for both the Poisson and binomial CDF approximations are presented in this section. Some preliminary definitions and two fundamental approximation theorems are discussed first in Section 2.1. The distributional lemmas used to motivate the Poisson and binomial approximations are then presented in Sections 2.2 and 2.3, along with some numerical examples.

2.1 Preliminary Definitions and Theorems

In order to quantify the precision of any CDF approximation, we first need to adopt a suitable definition of accuracy. Following Schader and Schmid (1989) and as discussed in Chang et al. (2008), define the "maximal absolute" (MABS) error of the binomial (or Poisson) CDF approximation as

$$MABS(F(k)|H(k)) = \max_{k\in\Omega} |F(k) - H(k)|, \qquad (3)$$

were $F(k)$ represents the exact CDF and $H(k)$ represents some proposed approximation to the CDF. Note that throughout the remainder of this manuscript, we will use the notation $MABS(Bin|\cdot)$ and $MABS(Poi|\cdot)$ to denote the MABS error for the various binomial and Poisson approximations, respectively.

The following two fundamental approximation theorems will play a critical role in the subsequent derivations.

Theorem 1 (The Wilson–Hilferty chi-square approximation). Let Q be distributed as a chi-square random variable (rv) with v degrees of freedom; that is, $Q \sim \chi^2_v$ and let $F_{\chi^2:v}(q)$ represent the corresponding chi-square CDF. Then $F_{\chi^2:v}(q) \approx \Phi((c - \mu)/\sigma)$ for $c = (q/v)^{1/3}$, $\mu = 1 - 2/(9v)$, and $\sigma = \sqrt{2/9v}$.

Theorem 2 (The Geary–Fieller approximation to a ratio of two independent normal rv's). Let W_1 and W_2 be distributed as two independent normal rv's with means and variances (μ_1, σ_1^2) and (μ_2, σ_2^2), respectively, and let $V = W_1/W_2$. Then $R = (V\mu_2 - \mu_1)/\sqrt{V^2\sigma_2^2 + \sigma_1^2}$ will approximately follow a standard normal distribution, provided that $\Pr(W_2 \le 0) \approx 0$.

Theorem 1 represents the well-known Wilson and Hilferty (1931) normal approximation to the chi-square distribution. This formula is known to produce very tight approximations to the chi-square CDF, even for χ^2 rv's exhibiting relatively small degrees of freedom. Theorem 2, which is discussed in detail in Fieller (1932), was first derived by Geary (1930) using a change-of-variable technique. Provided that the probability of W_2 being negative is negligible, this approximation works extremely well.

2.2 The Wilson–Hilferty Approximation to the Poisson CDF

The following lemma relating the Poisson and chi-square (gamma) distribution functions can be used to produce a tight, normal-CDF approximation to the Poisson CDF.

Lemma 1. Let $F_\lambda(k)$ represent the CDF of a Poisson(λ) distribution and let $F_{\chi^2:2(1+k)}(q)$ represent the CDF of a chi-square distribution with $2(1 + k)$ degrees of freedom. Then $F_\lambda(k) = 1 - F_{\chi^2:2(1+k)}(2\lambda)$.

Lemma 1 states the structural relationship between the Poisson(λ) and $\chi^2_{2(1+k)}$ distributions, a result that is discussed in many introductory graduate level probability and inference text books (e.g., see page 100 of Casella and Berger 2002). By combining Theorem 1 and Lemma 1 together, we can immediately deduce an attractive normal distribution approximation to the Poisson CDF; that is,

Corollary 1 (The Wilson–Hilferty normal approximation to the Poisson CDF). If $X \sim$ Poisson(λ) and $\lambda \ge 1$, then the CDF of X can be accurately approximated as $F_\lambda(k) \approx 1 - \Phi((c - \mu)/\sigma)$ for $c = (\lambda/(1+k))^{1/3}$, $\mu = 1 - 1/(9(1+k))$, and $\sigma = 1/(3\sqrt{1+k})$.

This approximation technique appears to have been originally suggested by Johnson and Kotz (1969), although we suspect that it had probably been in use long before these two authors published their seminal textbook on discrete distributions. Its derivation is fairly obvious, given the distributional relationship between the Poisson and gamma CDFs.

The approximation formula given in Corollary 1 is not typically presented in many statistical textbooks, although it probably should be. This Wilson–Hilferty (WH) normal-CDF approximation to the Poisson CDF is vastly superior to the limiting normal (LN) approximation, yet just as easy to compute. For example, for $F_{\lambda=10}(k)$ (the smallest value of λ for which the LN approximation is typically recommended) the absolute error in the WH approximations never exceeds 0.0005, while the LN approximations can produce errors as large as 0.0207.

Figure 1 presents the $MABS(Poi|WH)$ and $MABS(Poi|LN)$ error rates for $1 \le \lambda \le 25$; it is obvious from this plot that the WH approximation dominates the LN approximation. In general, the WH approximation to the Poisson CDF reduces the MABS error by about 2 orders of magnitude in comparison to the LN approximation for large λ values; at $\lambda = 100$ the $MABS(Poi|WH)$ error is just 0.00005, in contrast to an $MABS(Poi|LN)$ error rate of 0.00664. Likewise, the approximation remains applicable for very small λ values as well; even at $\lambda = 2$ the $MABS(Poi|WH)$ error is still only about 0.0025.

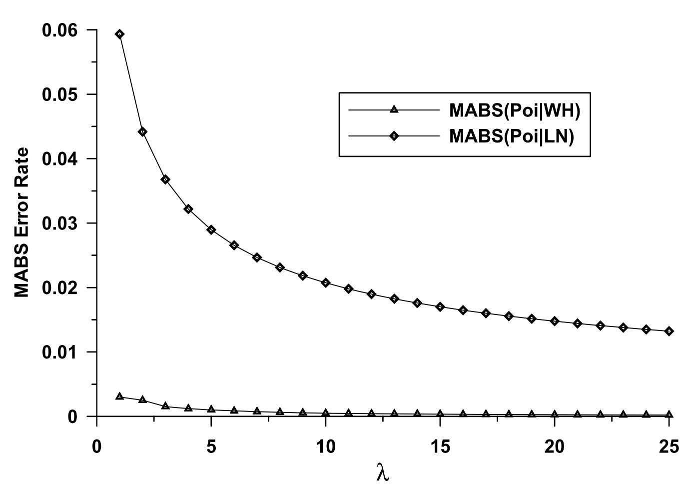

Figure 1. MABS error rates for the WH and LN approximations to the Poisson distribution, as a function of λ.

2.3 The Camp–Paulson Approximation to the Binomial CDF

The following two lemmas quantify the fundamental relationships between the binomial, beta and F CDFs; these relationships can be exploited to yield an accurate, normal-CDF approximation to the binomial distribution function.

Lemma 2. Let $F_{n,p}(k)$ represent the CDF of a Bin(n, p) distribution and let $F_{\beta:\nu,\omega}(x)$ represent the CDF of a Beta(ν, ω) distribution with shape parameters (ν, ω). Then $F_{n,p}(k) = 1 - F_{\beta:\nu,\omega}(p)$, where $\nu = k + 1$ and $\omega = n - k$.

Lemma 3. Let $F_{\beta:\nu/2,\omega/2}(x)$ represent the CDF of a $\beta(\nu/2, \omega/2)$ distribution with shape parameters $(\nu/2, \omega/2)$. Additionally, let $F_{F:\omega,\nu}(t)$ represent the CDF of a central F distribution with numerator and denominator degrees of freedom ω and ν, respectively. Then $F_{\beta:\nu/2,\omega/2}(\nu/(\nu + \omega t)) = 1 - F_{F:\omega,\nu}(t)$.

Lemma 2 establishes the fundamental CDF relationship between the binomial and beta *rv*'s. Like Lemma 1, this relationship can be established using integration by parts and is discussed in various distribution theory textbooks (i.e., see Johnson and Kotz 1969; page 63, equation 34). Lemma 3 defines the direct CDF relationship between the beta and F *rv*'s; this result can be readily established using a change-of-variable technique. Together, these two lemmas supply us with a direct link between the binomial and F distributions. More specifically, upon taking $\nu = 2(k + 1)$, $\omega = 2(n - k)$ and noting that $(\nu/(\nu + \omega t)) = p \Rightarrow t = [(k + 1)(1 - p)]/[(n - k)p]$, we can immediately establish that $F_{n,p}(k) = F_{F:2(n-k),2(k+1)}(t)$. Thus, any precise approximation to the F distribution can be immediately adapted to produce an equally precise approximation to the binomial distribution.

Paulson (1942) was one of the first statisticians to suggest a tight normal-CDF approximation to the F distribution, before comprehensive F tables were readily available. He noted that if T followed a central F distribution, then $T \sim (\chi^2_\omega/\omega)/(\chi^2_\nu/\nu)$ and thus the Wilson–Hilferty chi-square approximation (Theorem 1) suggested that $T^{1/3}$ should be approximately distributed as a ratio of two independent normal *rv*'s. Therefore, by making use of the Geary–Fieller approximation (Theorem 2), Paulson suggested taking $F_{F:\omega,\nu}(t) \approx \Phi((c-\mu)/\sigma)$ for $c = (1-b)t^{1/3}$, $\mu = 1 - a$, and $\sigma = \sqrt{bt^{2/3} + a}$, where $a = 2/(9\omega)$ and $b = 2/(9\nu)$, respectively. Paulson demonstrated that this approximation worked remarkably well, provided that ω and ν were ≥ 3. Camp (1951) appears to have been the first statistician to notice (or at least publish) that the Paulson normal-CDF approximation to the F distribution could be adapted to provide equally good approximations to the binomial CDF, by exploiting the distributional relationships discussed in Lemmas 2 and 3 above. His basic result is summarized in the following corollary:

Corollary 2 (The Camp–Paulson normal approximation to the binomial CDF). If $Y \sim$ Bin(n, p) and $n \geq 5$, then the CDF of Y can be accurately approximated as $F_{n,p}(k) \approx \Phi((c - \mu)/\sigma)$ for $c = (1 - b)r^{1/3}$, $\mu = 1 - a$, and $\sigma = \sqrt{br^{2/3} + a}$, where $a = 1/(9(n-k))$, $b = 1/(9(k+1))$, and $r = [(k+1)(1 - p)]/[(n - k)p]$, respectively.

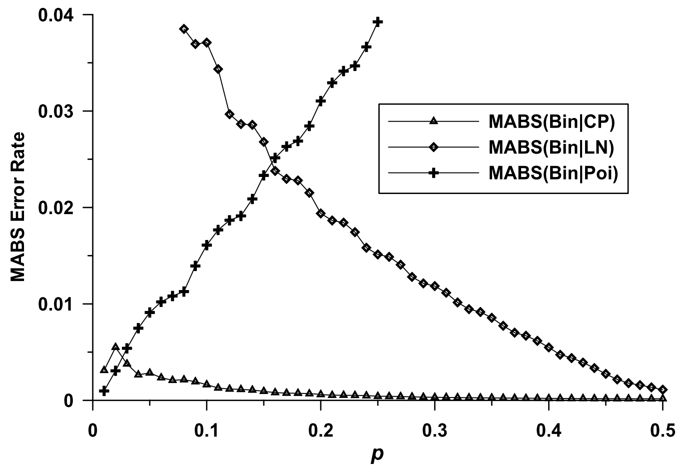

Figure 2. MABS error rates of the CP, LN, and Poisson approximations to the binomial ($n = 25$), as a function of p.

Like the Wilson–Hilferty approximation to the Poisson CDF, the Camp–Paulson (CP) approximation to the binomial CDF is vastly superior to the limiting normal approximation. Consider the typical example of approximating $F_{n=100,p=0.1}(k)$ for various integers k. In this example, the absolute error in the CP approximations never exceeds 0.0004, while the LN and Poisson approximations produce errors as large as 0.0175 and 0.0142, respectively.

For a sample size of $n = 25$, Figure 2 presents the *MABS*(*Bin*|*CP*), *MABS*(*Bin*|*LN*), and *MABS*(*Bin*|*Poi*) error rates for $0.01 \leq p \leq 0.5$. It is clear from this plot that the CP approximation dominates the LN approximation for all values of p. In contrast, the CP approximation does not entirely dominate the Poisson approximation; for values of p very close to 0 or 1 the Poisson approximation will still produce slightly better results.

Johnson and Kotz (1969) note that the error in the CP approximation can never exceed $0.007/\sqrt{np(1-p)}$ (a fairly conservative upper bound), while the error in the LN approximation can approach $0.140/\sqrt{np(1-p)}$. Peizer and Pratt (1968) and Pratt (1968) present some detailed asymptotic theory concerning the CP (and similar) approximation(s) and explain why such approximations dominate their LN counterparts. If a simple normal-CDF approximation formula is desired for the binomial distribution, the CP approximation should always be used in place of the LN approximation (for all values of p). It is also worthwhile to note that the CP approximation is more accurate than the skew normal approximation recently suggested by Chang et al. (2008) and does not require the numerical solution to either moment equations or the skew-standard CDF.

Finally, the fact that the (exact) Poisson approximation can still produce a smaller MABS value when p is near 0 or 1 suggests the need for a decision rule (for choosing between the CP and Poisson approximations). For large n, it can be numerically verified that $MABS(Bin|CP) \approx 0.003/\sqrt{np(1-p)}$. Likewise, the MABS error rate for the Poisson approximation becomes nearly independent of the sample size and about equal to $0.14p$. Equating these two limiting relationships together yields $0.000459 = np^3(1 - p)$. In turn, by taking $1 - p \approx 1$ we obtain the following decision rule. When computing an approximation

Table 1. The smallest value of p (and/or $1-p$) for which $MABS(Bin|CP) < MABS(Bin|Poi)$ for $n = 25, 50, 100, 250, 500$, and 1000. Decision rule values of $\tilde{p} = (0.0005/n)^{1/3}$ are shown in column 3.

n	Smallest computed p value (to 0.0005 precision)	Decision rule value $\tilde{p}$
25	0.0265	0.0271
50	0.0170	0.0215
100	0.0155	0.0171
250	0.0110	0.0126
500	0.0085	0.0100
1000	0.0065	0.0079

to the Bin(n, p) CDF for large n, if p or $1 - p < (0.0005/n)^{1/3}$ then use the Poisson approximation, else use the Camp–Paulson normal-CDF approximation presented in Corollary 2.

For selected n, Table 1 shows the numeric value of p for which $MABS(Bin|CP) < MABS(Bin|Poi)$; that is, the smallest value of p where the CP approximation outperforms the Poisson approximation. The value of p computed from the simple rule given above is also shown in Table 1. For $25 \leq n \leq 1000$, this simple decision rule works reasonably well, subject to a slight bias towards favoring the exact Poisson approximation.

3. DISCUSSION & CONCLUSION

In this article, we have reviewed and discussed two historical normal-CDF approximations to the Poisson and binomial distributions; that is, the Wilson–Hilferty and Camp–Paulson approximations. Both of these approximations reduce to standard normal formulas and produce very accurate estimates of the Poisson and binomial CDFs. Outside of a few extra algebraic steps, they are just as easy to apply as the classic limiting normal formulas, yet they perform remarkably better. Additionally, this improvement holds when one considers other error criteria as well, such as the average absolute (AABS) error rate. In the Poisson example discussed in Section 2.2 ($F_{\lambda=10}(k)$), one can easily verify that $AABS(Poi|WH) = 0.0002$ and $AABS(Poi|LN) = 0.0074$; similar results are obtained for the binomial example discussed in Section 2.3. Indeed, in any course where the limiting normal formulas are discussed in detail, it would certainly seem worthwhile to at least present these more precise approximations also. If nothing else, this would demonstrate to the typical student (who may only take one statistics course in his or her college career) that it is actually possible to derive accurate and simple approximations to the Poisson and binomial CDFs.

In a master's (or advanced undergraduate) probability and inference course, the derivation of these approximations presents a nice opportunity to introduce and study various distributional relationships. More specifically, these two approximations supply tangible examples of how some early statisticians exploited the distributional relationships between the gamma and Poisson distributions, and the binomial, beta, and F distributions to derive some rather elegant normal-CDF formulas. As noted in Section 2, these relationships can be readily established using either integration by parts or via change-in-variable techniques (two topics that are certainly worth emphasizing in a master's-level probability and inference course).

Finally, these two approximations also demonstrate the broad usefulness of the Wilson–Hilferty normal-CDF approximation to the chi-square distribution. The Wilson–Hilferty approximation technique clearly played a fundamental role in both of these derivations, in addition to accurate normal-CDF approximations for the beta, F and negative binomial distributions also (Johnson and Kotz 1969). Although the modern computing environment has obviously alleviated the necessity of this approximation, it is still both historically relevant and quite insightful from an instructional perspective. In fact, the derivation of these Poisson and binomial formulas present two nice examples of the role that the Wilson–Hilferty technique played in the mid-20th-century approximation literature.

[Received July 2008. Revised June 2009.]

REFERENCES

Camp, B. H. (1951), "Approximation to the Point Binomial," *Annals of Mathematical Statistics*, 22, 130–131.

Casella, G., and Berger, R. L. (2002), *Statistical Inference* (2nd ed.), Pacific Grove, CA: Duxbury.

Chang, C. H., Lin, J. J., Pal, N., and Chiang, M. C. (2008), "A Note on Improved Approximation of the Binomial Distribution by the Skew-Normal Distribution," *The American Statistician*, 62, 167–170.

Fieller, E. C. (1932), "The Distribution of the Index in a Normal Bivariate Population," *Biometrika*, 24, 428–440.

Geary, R. C. (1930), "The Frequency Distribution of the Quotient of Two Normal Variates," *Journal of the Royal Statistical Society*, 93, 442–446.

Johnson, N. L., and Kotz, S. (1969), *Discrete Distributions*, Boston, MA: Houghton Mifflin.

Paulson, E. (1942), "An Approximate Normalization of the Analysis of Variance Distribution," *Annals of Mathematical Statistics*, 13, 233–235.

Peizer, D. B., and Pratt, J. W. (1968), "A Normal Approximation for Binomial, F, beta, and Other Common, Related Tail Probabilities: I," *Journal of the American Statistical Association*, 63, 1416–1456.

Pratt, J. W. (1968), "A Normal Approximation for Binomial, F, beta, and Other Common, Related Tail Probabilities: II," *Journal of the American Statistical Association*, 63, 1457–1483.

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.

Wilson, E. B., and Kilferty, M. M. (1931), "The Distribution of Chi-Square," *Proceedings of the National Academy of Sciences of the USA*, 17, 684–688.