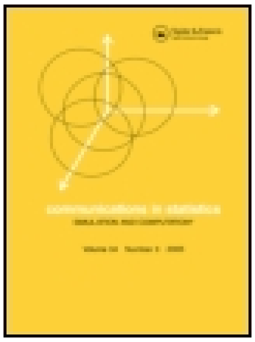


Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: http://www.tandfonline.com/loi/lssp20


# Critical review and comparison of continuity correction methods: The normal approximation to the binomial distribution


Takeshi Emura & Yu-Ting Liao



**To cite this article:** Takeshi Emura & Yu-Ting Liao (2017): Critical review and comparison of continuity correction methods: The normal approximation to the binomial distribution, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2017.1341527

**To link to this article:** http://dx.doi.org/10.1080/03610918.2017.1341527

Accepted author version posted online: 14 Jun 2017.
Published online: 14 Jun 2017.

COMMUNICATIONS IN STATISTICS—SIMULATION AND COMPUTATION®
2017, VOL. 0, NO. 0, 1–20
https://doi.org/10.1080/03610918.2017.1341527



# Critical review and comparison of continuity correction methods: The normal approximation to the binomial distribution

Takeshi Emura and Yu-Ting Liao

Graduate Institute of Statistics, National Central University, Taoyuan City, Taiwan

**ABSTRACT**
We provide a comprehensive and critical review of Yates' continuity correction for the normal approximation to the binomial distribution, in particular emphasizing its poor ability to approximate extreme tail probabilities. As an alternative method, we also review Cressie's finely tuned continuity correction. In addition, we demonstrate how Yates' continuity correction is used to improve the coverage probability of binomial confidence limits, and propose new confidence limits by applying Cressie's continuity correction. These continuity correction methods are numerically compared and illustrated by data examples arising from industry and medicine.

**ARTICLE HISTORY**
Received 29 September 2016
Accepted 5 June 2017

**KEYWORDS**
Central limit theorem; Confidence limit; Control chart; Statistical process control; Yates' correction

## 1. Introduction

In statistics, a continuity correction has long been employed when researchers try to approximate a discrete probability distribution by a continuous distribution. Yates (1934) introduced an ad-hoc rule of "adding 0.5" in order to improve the accuracy of the normal approximation to a discrete probability distribution, which is called "Yates' correction."

The investigation of the continuity correction has a fairly long history, originated from Feller (1968) and Cox (1970) who provided its mathematical foundations. Nowadays, continuity corrections are treated as an essential tool in many textbooks, e.g. for mathematical statistics [Chapter 3 of Casella and Berger (2002)], industrial statistics [Chapter 3 of Montgomery (2009)], and medical statistics [Chapter 9 of Everitt 2003]. In addition, the continuity correction has been widely used for handling discrete data, including the statistical analyses of defective counts in process control (Montgomery 2009), consecutive failure counts in system reliability (Makri and Psillakis 2011) and disease incident counts in clinic (Pradhan, Evans, and Banerjee 2016).

Aside from applications, the continuity correction has played a fundamental role in methodological and theoretical development. For instance, the continuity correction offers a method to improve the coverage probability of the confidence interval for the binomial probability (Blyth and Still 1983; Schader and Schmid 1990). Theory and methods to approximate significance probabilities and P-values in conditional and unconditional tests are discussed in Haber (1982).

**CONTACT** Takeshi Emura takeshiemura@gmail.com Graduate Institute of Statistics, National Central University, Taoyuan City, Taiwan.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lssp.

Researchers often use the normal approximation without paying much attention to the approximation error. A study of Duran and Albin (2009) gave the cases where a careful application of the normal approximation still produces a large error in terms of false alarm rate and the average run length. They showed that, even for a quite large sample size ($n = 1000$), the average run length of the resultant $np$-chart is far from the true value. A study of Brown, Cai, and DasGupta (2001) showed an unacceptably poor coverage performance of the Wald type confidence interval which is a straight forward application of the normal approximation to the binomial. Unfortunately, the study of the approximation error is very limited in the recent statistical literature, except for Hansen (2011) and Emura and Lin (2015) both of which focused on the normal approximation to the binomial.

Currently, most researchers employ Yates' ad-hoc correction of "adding 0.5." However, our numerical analyses will reveal that Yates' correction cannot improve the accuracy of the normal approximation in certain important cases. Besides, there exist more refined methods to correct the approximation which performs remarkably better than Yates' correction. Nevertheless, such methods have not been widely used in the literature. In this context, some critical review and numerical comparison for the continuity correction procedures are demanded.

In this paper, we provide a comprehensive and critical review of Yates' continuity correction to approximate the binomial distribution, in particular emphasizing its poor ability to approximate extreme tail probabilities. As an alternative method, we also review Cressie's finely tuned continuity correction (Cressie 1978). In addition, we demonstrate how Yates' continuity correction is applied to improve the coverage probability of binomial confidence limits. Here, we propose new continuity-corrected confidence limits by applying Cressie's finely tuned correction.

The paper is organized as follows. Section 2 reviews the background and introduces continuity correction methods. Section 3 considers an application to statistical process control and Section 4 considers an application to confidence limits. Section 5 concludes the paper.

## 2. Continuity correction for the binomial distribution

First, we review the backgrounds regarding the normal approximation to the binomial distribution. Next, we review Yates' continuity correction (Yates 1934) and Cressie's finely tuned continuity correction (Cressie 1978).

### 2.1. Normal approximation to the binomial distribution

Before discussing continuity correction, we introduce some mathematical notations and review the background behind the normal approximation. Let $X$ be a random variable with a probability function

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0,\ 1, \dots,\ n\},$$

where $n \geq 1$ is the sample size and $0 < p < 1$ is a binomial proportion. This distribution is the binomial distribution, denoted as $X \sim Bin(n,\ p)$. The cumulative distribution function (c.d.f.) is

$$F_X(k) = \Pr(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}.$$

One can write $X = \sum_{j=1}^{n} X_j$, where $X_1,\ X_2, \ldots,\ X_n$ are independent Bernoulli random variables with $\Pr(X_j = 0) = 1 - p$ and $\Pr(X_j = 1) = p$. The moment generating function (m.g.f.) of $X_j$ is $M_{X_j}(t) = E(e^{tX_j}) \ = pe^t + (1-p)$, which is defined for $\forall t$.

The study of the sum of Bernoulli random variables, $X = \sum_{j=1}^{n} X_j$, and its the normal approximation was initiated by De Moivre (1756). Nowadays, the central limit theorem (CLT) is a common approach to study the sum of independent and identically distributed random variables whose m.g.f. exists around $t = 0$ (Casella and Berger 2002). According to the CLT, one obtains

$$\lim_{n\to\infty} \Pr\left( \frac{X - np}{\sqrt{np(1-p)}} \le x \right) = \Phi(x), \quad \text{for } \forall x,$$

where $\Phi(\ x\ ) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$ is the c.d.f. of the standard normal distribution. Thus, if $n$ is large, one can obtain the "normal approximation to the binomial" as

$$F_X(k) = \Pr(X \le k) \ \approx \Phi\left( \frac{k - np}{\sqrt{np(1-p)}} \right).$$

The CLT only informs us that the approximation "$\approx$" is reasonably accurate for large $n$. In reality, however, the accuracy also depends on $p$. It is well known that the normal approximation is accurate if $np$ or $n(1-p)$ is large enough. This is because the value of $p$ very close to 0 or 1 produces a highly skewed binomial distribution, leading to an unrealistic approximation even for a very large $n$. Table 1 gives six existing criteria that guarantee the accurate performance of the normal approximation. Emura and Lin (2015) recommended the criterion "$(np > 10$ and $p \ge 0.1)$ or $(np > 15)$" in applications to statistical process control.

If these criteria are not met, the normal approximation should not be applied. For instance, one should not apply the normal approximation to $p = 0.0001$ and $n = 10000$ as none of the criteria in Table 1 is met. In this kind of small $p$, the Poisson approximation is suggested (Montgomery 2009).

**Table 1.** Six existing criteria for the normal approximation to the binomial.

| Criteria | Reference |
|---|---|
| C1. $np > 10$ and $p \ge 0.1$ | Wetherill and Brown (1991), Montgomery (2009) |
| C2. $np > 15$ | Johnson (2009) |
| C3. $np > 5$ and $n(1-p) > 5$ | Schader and Schmid (1989), Casella and Berger (2002), Ryan (2011) |
| C4. $np > 10$ and $n(1-p) > 10$ | Hahn and Meeker (1991) |
| C5. $n(1-p) > 9$ | Hald (1952, 1978), Schader and Schmid (1989) |
| C6. $(np > 10$ and $p \ge 0.1)$ or $(np > 15)$ | Emura and Lin (2015) |

### 2.2. *Yates' continuity correction*

A continuity correction is an adjustment that is made when a discrete distribution is approximated by a continuous distribution. Since the binomial distribution is discrete and the normal distribution is continuous, it is common to use the continuity correction in the approximation. The most popular continuity correction is Yates' correction (Yates 1934) defined as

$$F_X(k) \approx \Phi\left( \frac{k - np + 0.5}{\sqrt{np(1-p)}} \right) \quad k \in \{0,\ 1, \ldots,\ n\}.$$

One may consider a more "conservative correction" of adding 0.3 such that

$$F_X(k) \approx \Phi\left(\frac{k - np + 0.3}{\sqrt{np(1-p)}}\right), \quad k \in \{0,\ 1, \ldots,\ n\}.$$

To demonstrate the usefulness of Yates' continuity correction, we consider a simple example of $X \sim Bin(150,\ 0.1)$ and $k = 14$. Since $np = 15$ and $\sqrt{np(1-p)} = 3.6742$, one can approximate the binomial distribution as

$$F_X(14) = \Pr(X \leq 14) \approx \Phi\left(\frac{14 - 15}{3.6742}\right) = 0.3927.$$

Yates' continuity correction leads to

$$F_X(14) = \Pr(X \leq 14) \approx \Phi\left(\frac{14 - 15 + 0.5}{3.6742}\right) = 0.4459.$$

The true probability is

$$F_X(14) = \Pr(X \leq 14) = \sum_{k=0}^{14} \binom{150}{k} (0.1)^k (0.9)^{150-k} = 0.4602.$$

Clearly, Yates' correction gives a more precise approximation to the true probability. Figure 1 graphically explains why the continuity correction improves the accuracy in this example.

Regarding Yates' correction, an important remark should be mentioned.

**Remark I:** Yates' correction does not improve the accuracy for extreme tail probabilities where $F_X(k)$ is very close to either 0 or 1.

This remark was already mentioned by Yates (1934). However, this problem is neither described in many textbooks nor explored in journal articles. Since many real statistical problems are concerned about tail probabilities, Remark I may avoid the incorrect usage of the

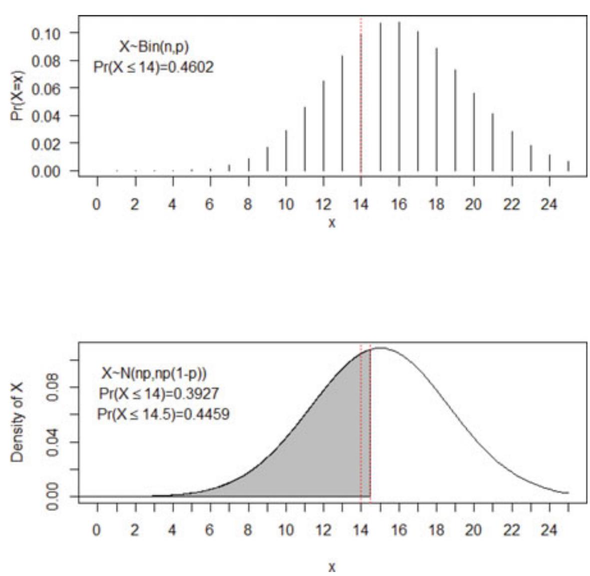


**Figure 1.** A normal approximation to the probability $F_X(14) = \Pr(X \leq 14)$ with and without a continuity correction under $n = 150$ and $p = 0.1$.

continuity correction by users. We will further demonstrate this problem through our subsequent numerical studies.

### 2.3. Finely tuned continuity correction (Cressie 1978)

A sophisticated continuity correction was proposed by Cressie (1978) for the normal approximation to the binomial distribution. The main argument of Cressie (1978) is to apply Stirling's formula to $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ and then take logarithms for the binomial probability $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. Then, a high-order Taylor series expansion is employed to derive the optimal correction rule. However, the mathematical arguments are extremely complicated. Accordingly, the finely tuned correction of Cressie (1978) is

$$F_X(k) \approx \Phi\left( \frac{k - np + d(k,\ p)}{\sqrt{np(1-p)}} \right), \quad k \in \{0,\ 1, \ldots,\ n\},$$

where the correction term is

$$d(\ k,\ \ p\ ) = 0.5 - (q - p)\left(\delta_{k-0.5}^2 - 1\ \right)/6,$$

where $q = 1 - p$, and $\delta_k = (\ k - np\ )/\sqrt{np(1-p)}$.

Our extensive investigations on the mathematical derivations of Cressie (1978) reveal his unclear mathematical arguments, especially about how the remainder terms are omitted. To complete the mathematical understanding of Cressie's derivations, additional research will be demanded. Nevertheless, the numerical performance of Cressie's correction is excellent.

## 3. Application to statistical process control

We demonstrate how the continuity correction is applied to problems on statistical process control.

### 3.1. np-control chart

The goal of $np$-control chart (or $np$-chart) is to control the fraction of nonconforming items produced in a factory [Chapter 10 of Wetherill and Brown (1991); Chap 7 of Montgomery (2009)]. Nowadays, the $np$- chart is one of the most important and fundamental control charts in statistical process control.

Let $X$ be the number of nonconforming items which follows $X \sim Bin(n,\ \ p)$, where $p$ is the fraction nonconforming. We consider a one-sided control chart, consisting of

$$\text{Center line} = np$$
$$UCL(\text{ upper control limit }) = np + 3\sqrt{np(1-p)}.$$

If $X \leq UCL$, then the process is declared "in-control." If $X > UCL$ then the process is declared "out-of-control." The value of $p$ is called "in-control value" which must be prespecified by an engineer. In some case, the value of $p$ may be an estimate from preliminary samples (Phase I samples). In any case, the value of $p$ is regarded as a known constant and is not estimated by $X$.

Accordingly, the in-control probability is defined as

$$P^* = \Pr(X \le UCL) = \Pr(X \le [\,UCL\,]) = \sum_{x \le [\,UCL\,]} \binom{n}{x} p^x (1-p)^{n-x}, \tag{1}$$

where $[\,UCL\,]$ is the largest integer not greater than $UCL$. By the CLT (Section 2.1), one can approximate the in-control probability as

$$P^* = \Pr(X \le [\,UCL\,]) = \Pr\left( \frac{X - np}{\sqrt{np(1-p)}} \le \frac{[\,UCL\,] - np}{\sqrt{np(1-p)}} \right) \approx \Phi\left( \frac{[\,UCL\,] - np}{\sqrt{np(1-p)}} \right).$$

In practice, the calculation of $P^*$ is frequently necessary, e.g., to compute the average run length and operating characteristic function (Wetherill and Brown 1991; Montgomery 2009).

### 3.2. Example of $n = 150$ and $p = 0.1$

For instance, we consider $np$-chart under $n = 150$ and $p = 0.1$. Then, $UCL = np + 3 \times \sqrt{np(1-p)} = 15 + 3 \times 3.6742 = 26.0227$ and hence, $[\,UCL\,] = 26$. Figure 2 shows an $np$-chart under this setting based on 50 replications. Engineers use the $np$-chart to see if the fraction of nonconforming is kept below the specified level of $p = 0.1$.

In the above setting, the in-control probability is

$$P^* = \sum_{x \le 26} \binom{150}{x} (0.1)^x (0.9)^{150-x} = 0.9981.$$

This means that 99.81% of data points fall below UCL (Figure 2). The value $P^*$ is approximated as

$$P^* = \Pr(X \le UCL) \approx \Phi\left( \frac{[\,UCL\,] - np + d}{\sqrt{np(1-p)}} \right),$$

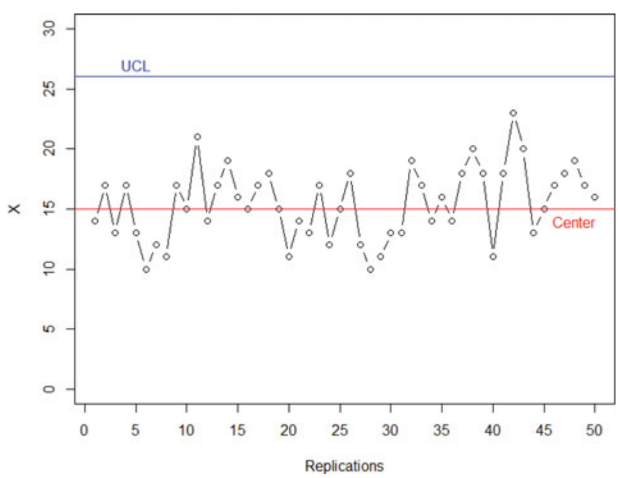


**Figure 2.** An example of $np$-control chart based on 50 replications under $n = 150$ and $p = 0.1$. The chart has the upper control limit $UCL = 26.0227$, Center line $= 15$, and the in-control probability $P^* = 0.9981$.

where $d$ is the correction term. The approximated in-control probability without continuity correction ( $d = 0$ ) is

$$\Phi\left(\frac{26-15}{3.6742}\right) = 0.9986, \quad (\text{Error} = 0.0005)\,.$$

Yates' continuity correction ( $d = 0.5$ ) results in

$$\Phi\left(\frac{26-15+0.5}{3.6742}\right) = 0.9991, \quad (\text{Error} = 0.001)$$

Hence, Yates' continuity correction does not improve the accuracy. Hence Yates' continuity correction should not be applied to the present case. This explains the poor performance of Yates' continuity correction for approximating extreme tail probabilities (0.9981 is very close to 1; see Remark I).

To consider Cressie's finely tuned continuity correction, we first calculate

$$\delta_{[\,UCL\,]-0.5} = \frac{[\,UCL\,]-0.5-np}{\sqrt{np(1-p)}} = \frac{26-0.5-15}{\sqrt{150\times 0.1\times(1-0.1)}} = 2.8577.$$

With the correction $d = 0.5-(q-p)(\delta^2_{[\,UCL\,]-0.5}-1)/6 = -0.4556$, we obtain the approximated in-control probability with Cressie's finely tuned continuity correction

$$\Phi\left(\frac{26-15-0.4556}{3.6742}\right) = 0.9979, \quad (\text{Error} = 0.0002)\,.$$

Hence, Cressie's correction greatly reduces the approximation error and shows the best performance among the three approximations.

### 3.3. Numerical assessment

Further numerical assessments were conducted to generalize the conclusion of the previous example of $n = 150$ and $p = 0.1$ (Section 3.2). We consider 19 different pairs:

$$\begin{aligned}(n,\,p) = {} & (800, 0.02),\ (550, 0.03),\ (400, 0.04),\ (350, 0.05),\\ & (300, 0.06),\ (250, 0.07),\ (200, 0.08), (200, 0.09),\ (150, 0.1),\\ & (100, 0.11),\ (90, 0.12),\ (80, 0.13), (80, 0.14),\ (70, 0.15),\\ & (70, 0.16),\ (60, 0.17),\ \ (60, 0.18),\ (60, 0.19),\ (60, 0.20).\end{aligned}$$

These choices were carefully made such that either ($np > 15$) or ($np > 10$ and $p \geq 0.1$) holds (Table 1). The first 8 pairs satisfy ($np > 15$) while the last 11 pairs satisfy ($np > 10$ and $p \geq 0.1$). See Appendix A for the detailed algorithm of choosing these ( $n,\ p$ ) pairs.

We define the error of approximating the true probability $P^*$ by

$$Err(\,n,\ p,\ [\,UCL\,]\,) = \left|\,P^* - \Phi\left(\frac{[\,UCL\,]-np+d}{\sqrt{np(1-p)}}\right)\right|,$$

where

$$P^* = \sum_{x\leq[UCL]}\binom{n}{x}p^x(1-p)^{n-x}, \quad UCL = np + 3\times\sqrt{np(1-p)}.$$

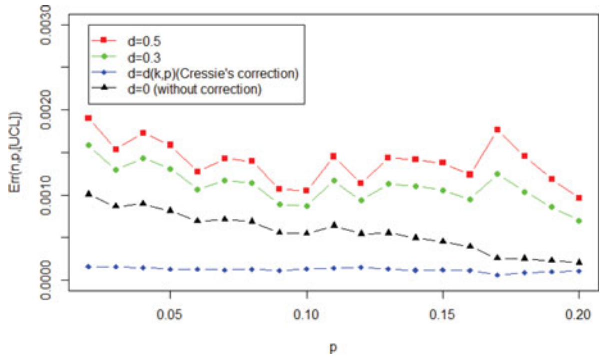


**Figure 3.** Comparison of four methods to approximate the true probability $P^*$. The error $Err(\ n,\ p,\ [\ UCL\ ]\ )$ is plotted against 19 different pairs of $(\ n,\ p\ )$'s.

In terms of the error, we compare the performance of the four continuity corrections: i) $d = 0.5$(Yates' correction), ii) $d = 0.5 - (q - p)(\delta^2_{[\ UCL\ ]-0.5} - 1)/6$ (Cressie's correction), iii) $d = 0.3$ (conservative correction), and iv) $d = 0$ (without correction).

Figure 3 summarizes the results. We find that using Cressie's correction provides the smallest error among the four corrections. Using Yates' correction ($d = 0.5$) results in the largest error. Accordingly, Yates' correction does not improve the usual normal approximation of $d = 0$ (without continuity correction). This result agrees with Remark I since we consider the extreme tails $P^* \approx 1$ in Equation (1). The conservative correction ($d = 0.3$) reduced the error compared to Yates' correction, but it is still poorer than the choice $d = 0$ (without continuity correction).

Numerical assessments in Figure 3 are specific to the application for statistical process control, where we use $k = [UCL]$. Now we perform additional numerical studies under more general statistical applications, where $\alpha = 0.05, 0.95$, and other extreme quantiles are of interest. Below, we examine the error of the normal approximation to $F_X(k) \approx \alpha$, where $\alpha = 0.9973$, 0.95, 0.05, and 0.0027. The error of the approximation is

$$Err(n,\ p,\ k) = \left| F_X(k) - \Phi\left( \frac{k - np + d}{\sqrt{np(1-p)}} \right) \right|.$$

For each $(\ n,\ p\ )$ and $\alpha$, we choose $k = [\ np + \Phi^{-1}(\alpha)\sqrt{np(1-p)}\ ]$.

The results are summarized in Figure 4. We find that using Cressie's finely tuned continuity correction performed excellently well, providing the smallest error among the four methods in majority of cases. One exception is the case of $\alpha = 0.95$, where $d = 0.3$ showed the smallest errors. The worst performance (largest error) is mostly achieved by Yates' correction ($d = 0.5$), showing no improvement over the usual normal approximation of $d = 0$ (without continuity correction). This is natural since Yates' correction should not be applied to the tail probability (Remark I). Another interesting finding is that the choice $d = 0.3$ is always better than $d = 0.5$. While there is no theoretical reason to use $d = 0.3$, it appears to give a good compromise between Yates' correction ($d = 0.5$) and the usual normal approximation ($d = 0$).

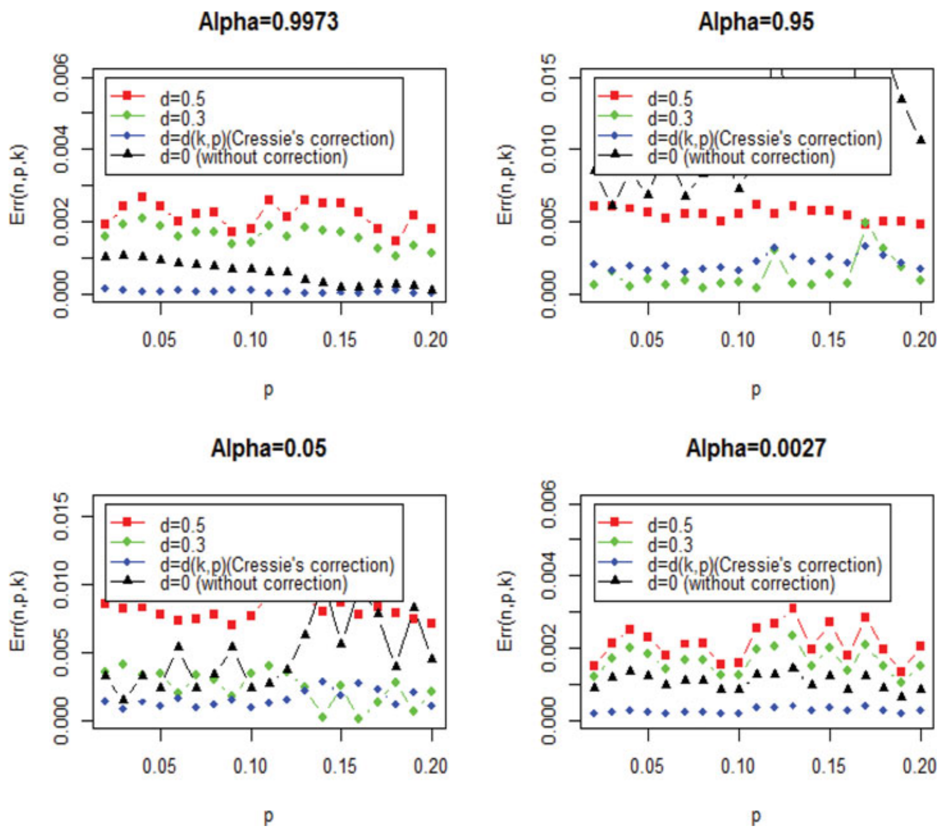


**Figure 4.** Comparison of four methods to approximate the true probability $F_X(k) \approx \alpha$. The error $Err(\ n,\ \ p,\ \ k\ )$ is plotted against 19 different pairs of $(\ n,\ \ p\ )$'s.

### *3.4. Data examples*

We compare three different continuity correction methods using two data examples: Data 1 is from Example 7.1 (pp. 292–293) and Data 2 is from Exercise 7.2 (p. 335) of Montgomery (2009). The two datasets are detailed in Appendices B and C.

Data 1 contains 30 replications of $n = 50$ orange juice cans, where cans were selected at half hour intervals (Appendix B). Data 1 contains total 347 nonconforming cans and the fraction nonconforming is estimated as

$$p = (12 + 15 + \cdots + 9 + 6)/(30 \times 50) = 347/1500 = 0.2313.$$

Then, $np = 11.567$, $UCL = np + 3 \times \sqrt{np(1-p)} = 20.512$, and $[\ UCL\ ] = 20$. Accordingly, the in-control probability is

$$P^* = \sum_{x \leq 20} \binom{50}{x} (0.2313)^x (1 - 0.2313)^{50-x} = 0.9976.$$

The error of approximating $P^* = 0.9976$ is

$$Err(n,\ \ p,\ \ [\ UCL\ ]) = \left| 0.9976 - \Phi\left( \frac{20 - 11.567 + d}{2.3097} \right) \right|.$$

If $d = 0.5$ (Yates' correction), we have $Err(n,\ p,\ [\ UCL\ ]) = 0.0010$ and if $d = 0.3$ , we have $Err(n,\ p,\ [\ UCL\ ]) = 0.0006$. To calculate Cressie's fine correction, we use

$$\delta_{[\ UCL\ ]-0.5} = \frac{[\ UCL\ ] - 0.5 - np}{\sqrt{np(1-p)}} = 2.6606,$$

$$d = 0.5 - \frac{(q-p)(\delta^2_{[UCL]-0.5} - 1)}{6} = -0.0444.$$

Then we have the error $Err(n,\ p,\ [\ UCL\ ]) = 0.0001$. Therefore, Cressie's finely continuity correction gives the best performance among the three continuity correction methods. Also, the correction $d = 0.3$ performs better than $d = 0.5$.

Data 2 consists of 20 replications of $n = 150$ titanium forgings for automobile turbocharger wheels (Appendix C). The overall fraction nonconforming is estimated as

$$p = (8 + 1 + 3 + \cdots + 3 + 0)/150/20 = 0.023.$$

Then, $np = 3.45$, $UCL = np + 3 \times \sqrt{np(1-p)} = 8.96$, and $[\ UCL\ ] = 8$. Accordingly, the in-control probability becomes

$$P^* = \sum_{x \le 8} \binom{150}{x} (0.023)^x (1 - 0.023)^{50-x} = 0.9918.$$

The error of approximating $P^* = 0.9918$ is

$$Err(n,\ p,\ [\ UCL\ ]) = \left| 0.9918 - \Phi\left( \frac{8 - 3.45 + d}{1.8359} \right) \right|.$$

If $d = 0.5$(Yates' correction), one has $Err(n,\ p,\ [\ UCL\ ]) = 0.0053$; if $d = 0.3$, one has $Err(n,\ p,\ [\ UCL\ ]) = 0.0041$. To calculate Cressie's finely continuity correction, we use

$$\delta_{[\ UCL\ ]-0.5} = \frac{[\ UCL\ ] - 0.5 - np}{\sqrt{np(1-p)}} = 2.206,$$

$$d = 0.5 - \frac{(q-p)(\delta^2_{[UCL]-0.5} - 1)}{6} = -0.1147.$$

Then one has $Err(n,\ p,\ [\ UCL\ ]) = 0.0004$. Therefore, Cressie's finely continuity correction gives the best performance among the three continuity correction methods. Also, the conservative correction ($d = 0.3$) performs slightly better than Yates' correction ($d = 0.5$).



## 4. Application to binomial confidence limits

We review methods to improve the accuracy of binomial confidence limits by applying continuity corrections. We follow an argument of Blyth and Still (1983) who applied Yates' continuity correction to the one-sided confidence limit or confidence interval for $p$. In our paper, we also propose new confidence limits by applying Cressie's finely tuned continuity correction. The new method is then compared with existing procedures by numerical analyses.

### 4.1. One-sided confidence limit with continuity correction

We explain how to apply a continuity correction to the one-sided confidence limit for $p$. While the two-sided interval (i.e., confidence interval) is more common than the one-sided limit, the latter gives us more transparent derivations of formulas, especially when applying Cressie's complicated correction method.

Let $X \sim Bin(n,\ p)$. Similar to Blyth and Still (1983), it follows that

$$
\begin{aligned}
&\Pr\left\{ \frac{(X+d)-np}{\sqrt{np(1-p)}} \le Z_\alpha \right\} \\
&= \Pr\left[ p^2 n\left(Z_\alpha^2 + n\right) - pn\left\{2(X+d) + Z_\alpha^2\right\} + (X+d)^2 \le 0 \right] \\
&= \Pr\left\{ p \le \frac{2n(X+d) + nZ_\alpha^2 + \sqrt{4(X+d)n^2 Z_\alpha^2 + n^2 Z_\alpha^4 - 4nZ_\alpha^2 (X+d)^2}}{2n(Z_\alpha^2+n)} \right\} \\
&= \Pr\left\{ p \le \frac{(X+d) + Z_\alpha^2/2 + Z_\alpha\sqrt{(X+d) - (X+d)^2/n + Z_\alpha^2/4}}{(Z_\alpha^2+n)} \right\},
\end{aligned}
$$

where $Z_\alpha = \Phi^{-1}(1-\alpha)$. By the CLT, an approximate "upper" confidence limit $U_d(X)$ with a continuity correction $d$ is

$$
U_d(X) = \frac{(X+d) + \frac{Z_\alpha^2}{2} + Z_\alpha\sqrt{(X+d) - \frac{(X+d)^2}{n} + \frac{Z_\alpha^2}{4}}}{(Z_\alpha^2+n)}
$$

for $X = 0, 1, \ldots, n-1$, and $U_d(n) = 1$. Blyth and Still (1983), Schader and Schmid (1990), and Casella and Berger (2002) considered $U_d(X)$ with $d = 0.5$ (Yates' correction).

#### 4.1.1. *New upper confidence limit by using the finely tuned correction*

To apply the finely tuned correction of Cressie (1978) to $U_d(X)$, one needs to calculate the correction term $d(k,\ p) = 0.5 - (q-p)(\delta_{k+0.5}^2 - 1)/6$, where $\delta_k = (k-np)/\sqrt{np(1-p)}$. If $(k,\ p)$ were known, one could use $U_{d(k,p)}(X)$ as the new upper confidence limit.

First, we consider how to compute $k$ given $p$. By the CLT,

$$
F_X(k) = \Pr(X \le k) \approx \Phi\left( \frac{k-np}{\sqrt{np(1-p)}} \right).
$$

The approximate value of $k$ is chosen such that

$$
\Phi\left( \frac{k-np}{\sqrt{np(1-p)}} \right) \approx \alpha.
$$

This yields

$$
k = \left[ np + \Phi^{-1}(\alpha)\sqrt{np(1-p)} \right].
$$

Next, we consider how to compute $p$ which is unknown. There are a number of different ways, such as the maximum likelihood estimate (MLE), minimax estimate, and Bayes estimators under various different priors. If we use the MLE $\hat{p} = X/n$, the upper confidence limit is

$$
U_{\hat{d}}(X) = \frac{(X+\hat{d}) + \frac{Z_\alpha^2}{2} + Z_\alpha\sqrt{(X+\hat{d}) - \frac{(X+\hat{d})^2}{n} + \frac{Z_\alpha^2}{4}}}{(Z_\alpha^2+n)}
$$

for $X = 0,\ 1,\ \ldots,\ n-1$, and $U_{\hat{d}}(n) = 1$, where $\hat{d} = d(\hat{k},\ \hat{p})$ and $\hat{k} = [\, n\hat{p} + \Phi^{-1}(\alpha)\sqrt{n\hat{p}(1-\hat{p})}\,]$.

### 4.1.2. Numerical assessments

The probability that the upper confidence limit covers the true $p$ is $C(n,\ p) = \Pr\{p \le U_d(X)\}$. To have a specified confidence level $(1-\alpha)$, it must satisfy the following condition

$$C(n,\ p) = \Pr\{p \le U_d(X)\} \ge 1-\alpha, \quad \forall p \in (0,1), \tag{2}$$

or equivalently $\inf_{p\in(0,1)} C(n,\ p) \ge 1-\alpha$. Equation (2) holds for $n \to \infty$ according to the CLT, but it is rarely true for finite samples. It has been known that Yates' continuity correction ($d = 0.5$) improves the accuracy of the confidence level in the sense of Equation (2); see Schader and Schmid (1990). It is our interest to examine if Cressie's finely tuned continuity correction further improves the performance relative to other correction methods.

We performed numerical analyses under the pairs of $(n,\ p)$'s with $(n = 50,\ p = 0.001,\ 0.002,\ 0.003, \ldots,\ 0.998, 0.999)$. For each $(n,\ p)$, we computed $C(n,\ p) = \Pr\{p \le U_d(X)\}$ under the confidence level $1-\alpha = 0.95$ using four different correction methods. Our evaluation criteria were whether the inequality $C(n,\ p) \ge 0.95$ holds for $p \in (0,1)$ and how $C(n,\ p)$ is close to 0.95.

Figure 5 shows the results. If the continuity correction is not employed ($d = 0$), the condition $C(n,\ p) \ge 0.95$ is violated in many cases in $0.001 \le p \le 0.999$. If Yates' correction is employed, the condition $C(n,\ p) \ge 0.95$ holds for $0.001 \le p \le 0.55$. If Cressie's correction is employed, the condition $C(n,\ p) \ge 0.95$ holds for $0.001 \le p \le 0.85$. Furthermore, Cressie's finely tuned correction gives the value of $C(n,\ p)$ closer to 0.95 than Yates' correction did. The performance of Cressie's correction method is almost unchanged even if $p$ is estimated by $\hat{p} = X/n$.

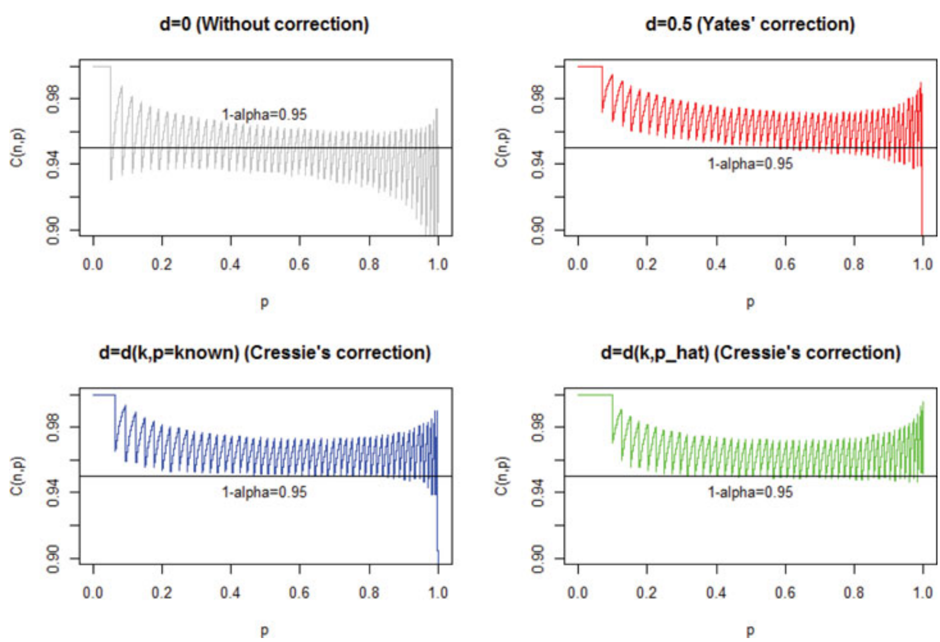


**Figure 5.** The coverage probability $C(n,\ p) = \Pr\{p \le U_d(X)\}$ for the four confidence limits under $(n = 50,\ p = 0.001,\ 0.002,\ 0.003, \ldots,\ 0.998, 0.999)$ and the confidence level $1-\alpha = 0.95$.

### 4.2. Confidence interval with continuity correction

We demonstrate how to apply a continuity correction to the confidence interval for $p$ based on $X \sim Bin(n,\ p)$. Following similar calculations to Section 4.1, the continuity-corrected confidence interval is $[\ L_c(X),\ U_d(X)\ ]$, where

$$L_c(X) = \frac{(\ X - c\ ) + \frac{Z^2_{\alpha/2}}{2} - Z_{\alpha/2}\sqrt{(\ X - c\ ) - \frac{(\ X-c\ )^2}{n} + \frac{Z^2_{\alpha/2}}{4}}}{\left(Z^2_{\alpha/2} + n\ \right)}$$

for $X = 1,\ 2,\ \ldots,\ n,$ and $L_c(0) = 0$, and

$$U_d(X) = \frac{(\ X + d\ ) + \frac{Z^2_{\alpha/2}}{2} + Z_{\alpha/2}\sqrt{(\ X + d\ ) - \frac{(\ X+d\ )^2}{n} + \frac{Z^2_{\alpha/2}}{4}}}{\left(Z^2_{\alpha/2} + n\ \right)}$$

for $X = 0,\ 1,\ \ldots,\ n-1,$ and $U_d(n) = 1$.

If $c = d = 0$ (without continuity correction), the confidence interval is called the Wilson interval (Wilson 1927; Brown, Cai, and DasGupta 2001). It is well-known that the Wilson interval (without continuity correction) has a superior coverage performance over the Wald interval (Brown, Cai, and DasGupta 2001; Casella and Berger 2002; Pradhan, Evans and Banerjee 2016).

Blyth and Still (1983), Schader and Schmid (1990), and Casella and Berger (2002) considered the interval $[\ L_c(X),\ U_d(X)\ ]$ with $c = d = 0.5$ (Yates' correction). The resultant interval will be called the Wilson interval with Yates' correction.

#### 4.2.1. New confidence interval by using the finely tuned correction

Using similar mathematical arguments to the one-sided case (Section 4.1), we propose to apply the finely tuned correction of Cressie (1978) to the confidence interval $[\ L_c(X),\ U_d(X)\ ]$. We suggest replacing the correction terms $c$ and $d$ with

$$c(\ k_L,\ p\ ) = 0.5 + (q - p)\left(\delta^2_{k_L - 0.5} - 1\ \right)/6,$$
$$d(\ k_U, p\ ) = 0.5 - (q - p)\left(\delta^2_{k_U + 0.5} - 1\ \right)/6,$$

where $q = 1 - p$, $\delta_k = (\ k - np\ )/\sqrt{np(1-p)}$,

$$k_L = \left[\ np + \Phi^{-1}(1 - \alpha/2)\sqrt{np(1-p)}\ \right],\ k_U = \left[\ np + \Phi^{-1}(\alpha/2)\sqrt{np(1-p)}\ \right].$$

The estimates $\hat{k}_L$ and $\hat{k}_U$ are obtained by replacing $p$ by $\hat{p} = X/n$. Consequently, the confidence interval is $[\ L_{\hat{c}}(X),\ U_{\hat{d}}(X)\ ]$, where $\hat{c} = c(\ \hat{k}_L,\ \hat{p}\ )$, and $\hat{d} = d(\ \hat{k}_U,\ \hat{p}\ )$. We will also consider the shrinkage estimate $\tilde{p} = (X + 2)/(n + 4)$ due to Agresti and Coull (1998).

#### 4.2.2. Numerical assessments

The probability that the confidence interval covers the true $p$ is $C(n,\ p) = \Pr\{L_c(X) \le p \le U_d(X)\}$. To have a specified confidence level $(1 - \alpha)$, it must satisfy

$$C(n,\ p) = \Pr\{L_c(X) \le p \le U_d(X)\} \ge 1 - \alpha, \quad \forall p \in (0, 1), \tag{3}$$

or equivalently $\inf_{p \in (0,1)} C(n,\ p) \ge 1 - \alpha$. However, Equation (3) is rarely true for finite samples (e.g., Brown, Cai, and DasGupta 2001). It has been known that Yates' continuity correction improves the accuracy of the confidence level in the sense of Equation (3); see Blyth and Still (1983), Schader and Schmid (1990), and Casella and Berger (2002). It is our interest to

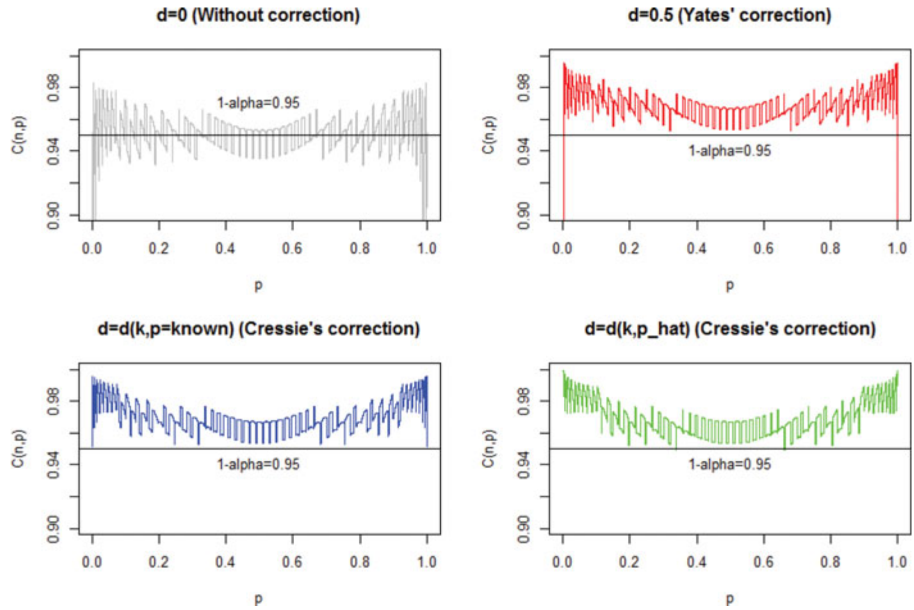


**Figure 6.** The coverage probability $C(n,\ p) = \Pr\{L_c(X) \le p \le U_d(X)\}$ for the 95% confidence intervals under ( $n = 50,\ p = 0.001,\ 0.002,\ 0.003, \ldots,\ 0.998, 0.999$). The four panels correspond to the Wilson intervals with four different continuity correction terms. The bottom-left panel is Cressie's correction with $p$ assumed known while the bottom-right panel is Cressie's correction with $p$ estimated by $\hat{p} = X/n$.

examine if Cressie's finely tuned continuity correction further improves the performance over Yates' correction.

We performed numerical analyses under ( $n = 50,\ p = 0.001,\ 0.002,\ 0.003, \ldots,\ 0.998,$ $0.999$). For each ( $n,\ p$ ), we computed $C(n,\ p) = \Pr\{L_c(X) \le p \le U_d(X)\}$ under the confidence level $1 - \alpha = 0.95$ using four different correction methods. Our evaluation criteria were whether the inequality $C(n,\ p) \ge 0.95$ holds for $p \in (0, 1)$ and how $C(n,\ p)$ is close to 0.95.

Figure 6 shows the results. If the continuity correction is not employed ($d = 0$), the condition $C(n,\ p) \ge 0.95$ is violated in many cases in $0.001 \le p \le 0.999$. If Yates' correction is employed, the condition $C(n,\ p) \ge 0.95$ holds for all cases in $0.002 \le p \le 0.998$, but does not for $p = 0.001$ and $p = 0.999$. If Cressie's correction is employed, the condition $C(n,\ p) \ge 0.95$ holds everywhere for $0.001 \le p \le 0.999$. However, Yates' correction and Cressie's correction have similar values of $C(n,\ p)$. The performance of Cressie's correction is less changed when $p$ is estimated by $\hat{p} = X/n$.

In the literature, there exist other methods to construct the binomial confidence interval. Hence, it would be interesting to see how these methods perform against the method based on Cressie's correction. We consider four methods:

(i) **Wald interval**: [ $L_W(X),\ L_W(X)$ ], where

$$L_W(X) = \hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad U_W(X) = \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} = \frac{X}{n}.$$

(ii) **Agresti-Coull interval** (Agresti and Coull 1998): [ $L_{AC}(X),\ L_{AC}(X)$ ], where

$$L_{AC}(X) = \tilde{p} - Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}, \quad U_{AC}(X) = \tilde{p} + Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}, \quad \tilde{p} = \frac{X+2}{n+4}.$$

This is the Wald interval with $\hat{p}$ replaced by its shrinkage estimator $\tilde{p}$ (Agresti and Coull 1998).

(iii) **Jeffreys interval**: $[\ L_J(X),\ L_J(X)\ ]$, where

$$L_J(X) = B^{-1}(\alpha/2;\ X + 0.5, n - X + 0.5),$$
$$U_J(X) = B^{-1}(1 - \alpha/2;\ X + 0.5, n - X + 0.5),$$

where $B^{-1}(p\ ;\ a,\ b)$ is the $p$-th quantile of a Beta distribution with parameters $a$ and $b$. This quantile is based on the posterior distribution of $p$ under the Jeffreys prior (the Beta distribution with $a = b = 0.5$).

(iv) **Wilson interval with Cressie's correction term estimated by Agresti-Coull estimate**:

$$[\ L_{\tilde{c}}(X),\ U_{\tilde{d}}(X)\ ], \quad \text{where } \tilde{c} = c(\ \tilde{k}_L,\ \tilde{p}\ ), \quad \tilde{d} = d(\ \tilde{k}_U,\ \tilde{p}\ ), \quad \tilde{p} = \frac{X+2}{n+4}.$$

This is the proposed interval using the shrinkage estimate $\tilde{p}$ (Agresti and Coull 1998).

Figure 7 shows the results. The Wald interval showed the worst performance due to the considerable under-coverage, $C(n,\ p) << 0.95$, for most cases of $0.001 \leq p \leq 0.999$. The Agresti and Coull method substantially improved upon the Wald interval and were competitive with the proposed interval (Wilson interval with Cressie's correction). The Jefferys interval also performed considerably better than the Wald interval, but violated the condition $C(n,\ p) \geq 0.95$ in many values of $p$. Nevertheless, the Jefferys interval had the smallest error $|C(n,\ p) - 0.95|$ on average (not shown). In summary, the proposed interval had competitive performance with the Agresti and Coull interval and exhibited a good control for the confidence level.

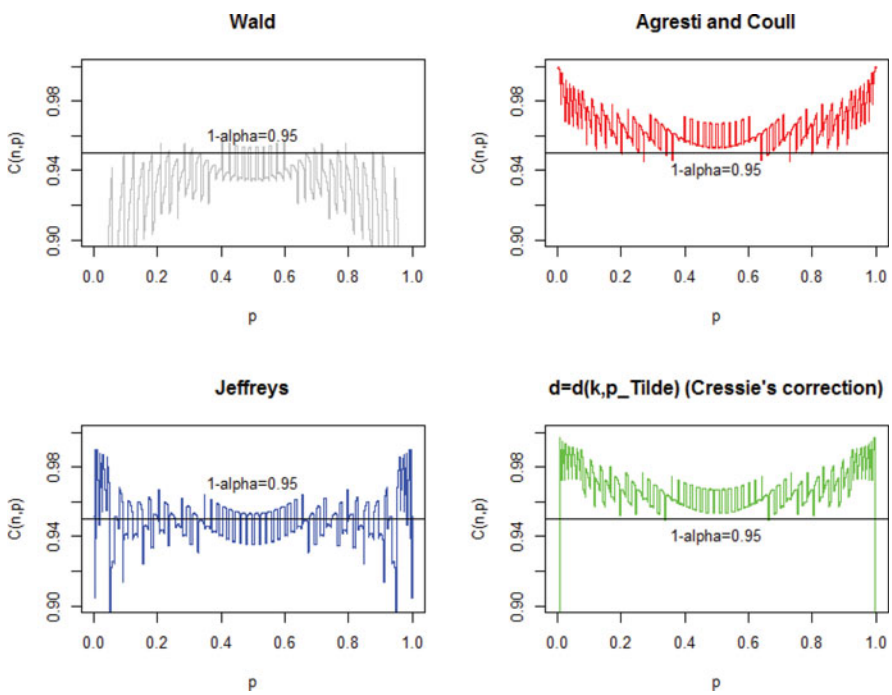


**Figure 7.** The coverage probability $C(n,\ p) = \Pr\{L_c(X) \leq p \leq U_d(X)\}$ for the 95% confidence intervals under ( $n = 50,\ p = 0.001,\ 0.002,\ 0.003, \ldots,\ 0.998, 0.999$). The four panels correspond to the Wald interval (upper-left), the Agresti and Coull interval (upper-right), the Jeffreys interval (bottom-left), and the Wilson interval with Cressie's correction with $p$ estimated by $\tilde{p} = (X+2)/(n+4)$ (bottom-right).

**Table 2.** 95% confidence intervals calculated for children with cardiac arrest.

| | Standard-dose group ($n = 34$) $X = 7$, $\hat{p} = X/n = 0.2059$ | | High-dose group ($n = 34$) $X = 1$, $\hat{p} = X/n = 0.0294$ | |
|---|---|---|---|---|
| Method | 95% Conf. interval | Length | 95% Conf. interval | Length |
| Wald | (0.0700, 0.3418) | 0.2718 | (−0.0274, 0.0862) | 0.1136 |
| Agresti and Coull | (0.0939, 0.3797) | 0.2858 | (−0.0117, 0.1696) | 0.1813 |
| Jeffreys | (0.0971, 0.3418) | 0.2447 | (0.0032, 0.0862) | 0.0830 |
| Wilson ($c = d = 0$) | (0.1035, 0.3680) | 0.2645 | (0.0052, 0.1492) | 0.1439 |
| Wilson ($c = d = 0.5$) | (0.0934, 0.3841) | 0.2907 | (0.0015, 0.1705) | 0.1690 |
| Wilson ($c = \hat{c},\ d = \hat{d}$)* | (0.0894, 0.3776) | 0.2883 | (0.0007, 0.1632) | 0.1625 |
| Wilson ($c = \tilde{c},\ d = \tilde{d}$)** | (0.0897, 0.3782) | 0.2885 | (0.0005, 0.1605) | 0.1600 |

*Cressie's correction $\hat{c} = c(\hat{k}_L,\ \hat{p})$ and $\hat{d} = d(\hat{k}_U,\ \hat{p})$, where $\hat{p} = X/n$

**Cressie's correction $\tilde{c} = c(\tilde{k}_L,\ \tilde{p})$ and $\tilde{d} = d(\tilde{k}_U,\ \tilde{p})$, where $\tilde{p} = (X + 2)/(n + 4)$

### *4.3. Data example*

We apply the seven methods for calculating the confidence interval (Section 4.2) to the data on children with cardiac arrest (Perondi et al. 2004). We used the data summarized in Table 2 of Fagerland, Lydersen, Laake (2015) that consists of two treatment groups: standard-dose group ($n = 34$) and high-dose group ($n = 34$). The standard group has $X = 7$ survivors with the success rate $\hat{p} = X/n = 0.2059$ while the high dose group has $X = 1$ survivor with the success rate $\hat{p} = X/n = 0.0294$. We will calculate the confidence intervals for the success rates in each group.

The 95% confidence intervals for the seven methods are given in Table 2.

In the standard-dose group, all the seven methods yielded similar confidence intervals. The Jeffreys interval gave the shortest interval length, and the Wilson interval (without correction) was the second shortest. This may be expected as these two intervals occasionally lead to under-coverage (i.e., $C(n,\ p) < 0.95$ in Figures 6 and 7). The Agresti and Coull interval was very similar to the three Wilson intervals (with three different continuity corrections).

In the high-dose group, the seven methods yielded remarkably different behavior, except that the Jeffreys interval again achieved the shortest interval length. First, the Wald interval and the Agresti and Coull interval exhibited negative values in the lower limit. This disadvantage did not appear in the four Wilson intervals. The Wilson interval without continuity correction is somewhat different from the Wilson intervals with continuity corrections. This highlights the strong influence of a continuity correction in the high-dose group. If the negative lower limit of the Agresti and Coull interval is simply regarded as 0, the resultant interval is very similar to the Wilson interval with continuity corrections.

In conclusion, the three Wilson intervals with three different continuity corrections gave similar results on the confidence interval of the success rates in both groups. Hence, they may be more trusted than other intervals that exhibit some extreme behavior. Perhaps, the two Wilson intervals with Cressie's corrections (the last two in Table 2) may be recommended as they had shorter interval lengths than the Wilson interval with Yates' correction.

## 5. Conclusion

Statisticians have long recognized Yates' continuity correction to be useful and convenient for approximating the discrete probability distribution. This paper shows that Yates' correction does not improve the accuracy for approximating extreme tail probabilities. While this issue was initially observed by Yates (1934), it has not been seriously discussed in the literature. As a promising solution to the problem, we review Cressie's finely tuned continuity correction

(Cressie 1978) that produced very accurate approximation to extreme tail probabilities in our numerical analysis. In addition to the normal approximation, we also review how these continuity corrections are applied to improve the coverage performance of the confidence limits for binomial proportion.

A numerical study is performed, comparing Yates' correction, Cressie's finely tuned correction, and a conservative correction of "adding 0.3." We try our numerical study to be practical by considering applications to statistical process control with the in-control probability of the $np$-chart and applications to statistical inference with tail probabilities corresponding to $\alpha =$ 0.9973, 0.95, 0.05, and 0.0027. In majority of cases, Yates' correction gave the worst performance while Cressie's finely tuned correction achieved the best performance. The failure of Yates' correction is due to our settings that focused on the normal approximation to extreme tail probabilities (close to 0 or 1). Interestingly, the conservative correction performed better than Yates' correction in these setting. However, we have no theoretical justification of using the value of 0.3.

A novel contribution of this paper is a newly proposed continuity-corrected confidence limit (interval) for the binomial proportion in Sections 4.1.1 (Section 4.2.1). The new confidence limit is derived by applying Cressie's continuity correction to the Wilson interval of Blyth and Still (1983). Unlike the setting of statistical process control where $p$ is known, Cressie's correction and Yates' correction yield similar performance on the coverage probabilities of their corrected confidence limits. Both of them appear to work well on the Wilson interval. This also reflects the difference between "how to reduce the error of a normal approximation to the binomial" and "how to improve the coverage performance of binomial confidence intervals."

## Acknowledgments

We thank the reviewer for his/her constructive suggestions about confidence intervals for binomial proportion. We also thank my three colleagues (Professors Cheng-Der Fuh, Huei-Wen Teng, and Li-Hsien Sun) for their comments on the earlier version of our manuscript and for Professor Nan-Cheng Su for his comments when our paper was presented at the 25th South Taiwan Statistics Conference. Our work is supported by the Taiwan government (MOST 103-2118-M-008-MY2; MOST 105-2118-M-008-003-MY2).

## Author contribution

- Emura T: Designed and supervised the study. Wrote and revised the manuscript (all sections). Conducted numerical analyses (Section 4).
- Liao YT: Wrote and revised the manuscript (all sections). Conducted numerical analyses (Sections 2 and 3).

## Appendix A: Choice of $(n, p)$

We explain how we chose pairs of $(n, p)$ for our numerical analyses. We follow the rule ($np >$ 10 and $p \geq 0.1$ ) or ($np > 15$ ) given by Emura and Lin (2015) to choose the value of $(n, p)$ from $24 \times 20 = 480$ pairs of

$$n = (\underbrace{10,\ 20, \ldots,\ 90,\ 100,\ 150, \ldots,\ 750,\ 800}_{\times 24}),$$

$$p = (\underbrace{0.01,\ 0.02, \ldots,\ 0.19,\ 0.20}_{\times 20}).$$

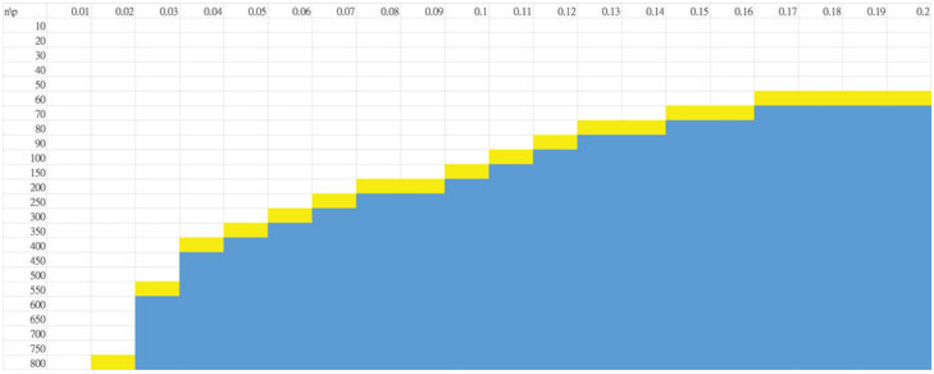


**Figure A.** The areas satisfying ($np > 10$ and $p \geq 0.1$ ) or ($np > 15$ ).

In Figure A, we display
(blue) satisfy the rule ($np > 10$ and $p \geq 0.1$ )or ($np > 15$ )
(yellow) satisfy the rule and it is the boundary of the rule ($np > 10$ and $p \geq 0.1$ )or ($np > 15$ )

The boundary cases (yellow color in Figure A) result in

$$\begin{aligned}(n, p) = &(800, 0.02),\ (550, 0.03),\ (400, 0.04),\ (350, 0.05),\\ &(300, 0.06),\ (250, 0.07),\ (200, 0.08),\ (200, 0.09),\ (150, 0.1),\\ &(100, 0.11),\ (90, 0.12),\ (80, 0.13),\ (80, 0.14),\ (70, 0.15),\\ &(70, 0.16),\ (60, 0.17),\ (60, 0.18),\ (60, 0.19),\ (60, 0.20).\end{aligned}$$

## Appendix B: Data 1 (orange juice cans)

The data come from Example 7.1 (pp. 292–293) of Montgomery (2009). In 6-oz cans, frozen orange juice concentrate is packed. By inspection of a can, manufacturers determine whether the can leaks either on the side seam or around the bottom joint. Table B shows the number of leaks (nonconformings) in $n = 50$ cans during 30 replications.

**Table B.** Data 1 (orange juice cans).

| Replication | The number of nonconforming cans | Replication | The number of nonconforming cans |
|---|---|---|---|
| 1 | 12 | 16 | 8 |
| 2 | 15 | 17 | 10 |
| 3 | 8 | 18 | 5 |
| 4 | 10 | 19 | 13 |
| 5 | 4 | 20 | 11 |
| 6 | 7 | 21 | 20 |
| 7 | 16 | 22 | 18 |
| 8 | 9 | 23 | 24 |
| 9 | 14 | 24 | 15 |
| 10 | 10 | 25 | 9 |
| 11 | 5 | 26 | 12 |
| 12 | 6 | 27 | 7 |
| 13 | 17 | 28 | 13 |
| 14 | 12 | 29 | 9 |
| 15 | 22 | 30 | 6 |

## Appendix C: Data 2 (titanium forgings for automobile turbocharger wheels)

The data come from Exercise 7.2 (p. 335) of Montgomery (2009). Manufactures examine titanium forgings for automobile turbocharger wheels in a set of $n = 150$ samples. Table C shows the number of nonconforming switches observed during 20 days.

**Table C.** Data 2 (titanium forgings for automobile turbocharger wheels).

| Replication (day) | The number of nonconformings | Replication (day) | The number of nonconformings |
|---|---|---|---|
| 1 | 8 | 11 | 6 |
| 2 | 1 | 12 | 0 |
| 3 | 3 | 13 | 4 |
| 4 | 0 | 14 | 0 |
| 5 | 2 | 15 | 3 |
| 6 | 4 | 16 | 1 |
| 7 | 0 | 17 | 15 |
| 8 | 1 | 18 | 2 |
| 9 | 10 | 19 | 3 |
| 10 | 6 | 20 | 0 |

## References


Agresti, A., Coull, B. A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician* 52(2):119–126.

Blyth, C. R., Still, H. A. (1983). Binomial confidence intervals. *Journal of the American Statistical Association* 81:108–116.

Brown, L. D., Cai, T. T., DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science* 16 (2):101–117.

Casella, G., Berger, R. L. (2002). *Statistical Inference*. 2rd ed. Australia: Duxbury Press.

Cox, D. R. (1970). The continuity correction. *Biometrika* 57:217–219.

Cressie, N. (1978). A finely tuned continuity correction. *Ann. Inst. Statist. Math* 30:435–442.

De Moivre, A. (1756). *The doctrine of chances*. 3rd ed. New York: Chelsea. Third edition, reprinted in 1967. First edition 1718.

Duran, R. I., Albin, S. L. (2009). Monitoring a fraction with easy and reliable settings of the false alarm rate. *Quality and Reliability Engineering International* 25(8):1029–1043.

Emura, T., Lin, Y. S. (2015). A comparison of normal approximation rules for attribute control charts. *Quality and Reliability Engineering International* 31 (No.3):411–418.

Everitt, B. (2003). *Modern medical statistics: A practical guide*. New York: Wiley.

Fagerland, M. W., Lydersen, S., Laake, P. (2015). Recommended confidence intervals for two independent binomial proportions. *Statistical Methods in Medical Research* 24(2):224–254.

Feller, W. 1968. *An introduction to probability theory and its applications*. volume I, 3rd ed. New York: John Wiley & Sons, Inc.

Haber, M. (1982). The continuity correction and statistical testing. *International Statistical Review* 50:135–144.

Hahn, G. J., Meeker, W. Q. (1991). *Statistical intervals: A guide for practitioners*. New York: Wiley.

Hald, A. (1952). *Statistical theory with engineering applications*. New York: Wiley.

Hald, A. (1978). *Statistical theory of sampling inspection. Part 2*. Institute of Mathematical Statistics, University of Copenhagen.

Hansen, P. (2011). Approximating the binomial distribution by the normal distribution - error and accuracy, *U.U.D.M. Project Report* 2011:18.

Johnson, R. A. (2009). *Statistics: Principles and Methods*. 6th ed. New York: Wiley.

Makri, F. S., Psillakis, Z. M. (2011). On runs of length exceeding a threshold: Normal approximation. *Statistical Papers* 52(3):531–51.

Montgomery, D. C. (2009). *Statistical quality control: A modern introduction*. 6rd ed. New York: John Wiley & Sons.

Perondi, M. B. M., Reis, A. G., Paiva, E. F., Nadkarni, V. M., Berg, R. A. 2004. A comparison of high-dose and standard-dose epinephrine in children with cardiac arrest. *New England Journal of Medicine* 350(17):1722–1730.

Pradhan, V., J. C. Evans, Banerjee, T. 2016. Binomial confidence intervals for testing non-inferiority or superiority: a practitioner's dilemma. *Statistical Methods in Medical Research* 25 (4):1707–1717.

Ryan, T. P. (2011). *Statistical methods for quality improvement*. 3rd ed. Wiley Series in Probability and Statistics. New York: John Wiley & Sons.

Schader, M., Schmid, F. (1989). Two rules of thumb for the approximation of the binomial distribution by the normal distribution. *American Statisticians* 43(1):23–24.

Schader, M., Schmid, F. (1990). Charting small sample characteristics of asymptotic confidence intervals for the binomial parameter p. *Statistical Papers* 31:251–64.

Wetherill, G. B., Brown, D. W. (1991). *Statistical process control: Theory and practice*. Chapman and Hall: London.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association* 22(158):209–12.

Yates, F. (1934). Contingency tables involving small numbers, and the $\chi^2$ test. *Journal of the Royal Statistical Society* 1(2):217–35.