Journal of Statistical Planning and Inference 141 (2011) 1709–1718

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Improved closed-form prediction intervals for binomial and Poisson distributions

K. Krishnamoorthy [a,*], Jie Peng [b]

[a] Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504, United States
[b] Department of Finance, Economics and Decision Science, St. Ambrose University, Davenport, IA 52803, United States

ARTICLE INFO

Article history:
Received 1 May 2010
Received in revised form
16 November 2010
Accepted 22 November 2010
Available online 1 December 2010

Keywords:
Clopper–Pearson approach
Conditional approach
Coverage probability
Hypergeometric distribution

ABSTRACT

The problems of constructing prediction intervals for the binomial and Poisson distributions are considered. Available approximate, exact and conditional methods for both distributions are reviewed and compared. Simple approximate prediction intervals based on the joint distribution of the past samples and the future sample are proposed. Exact coverage studies and expected widths of prediction intervals show that the new prediction intervals are comparable to or better than the available ones in most cases. The methods are illustrated using two practical examples.

1. Introduction

In many practical situations, one needs to predict the values of a future random variable based on the past and currently available samples. Extensive literature is available for constructing prediction intervals (PIs) for various continuous probability distributions and other continuous models such as linear regression and one-way random models. Applications of prediction intervals (PIs) based on continuous distributions are well-known; for examples, the prediction intervals based on gamma distributions are often used in environment monitoring (Gibbons, 1987), and normal based PIs are used in monitoring and control problems (Davis and McNichols, 1987). Other applications and examples can be found in the book by Aitchison and Dunsmore (1980). Compared to continuous distributions, results on constructing PIs for discrete distributions are very limited. Prediction intervals for a discrete distribution are used to predict the number of events that may occur in the future. For example, a manufacturer maybe interested in assessing the number of defective items in the future production process based on available samples. Faulkenberry (1973) provided an illustrative example in which the number of breakdowns of a system in a year follows a Poisson distribution, and the objective is to predict the number of breakdowns in a future year based on available samples. Bain and Patel (1993) have noted a situation where it is desired to construct binomial prediction intervals (see Section 4).

The prediction problem that we will address concerns two independent binomial samples with the same "success probability" p. Given that X successes are observed in n independent Bernoulli trials, we like to predict the number of successes Y in another m independent Bernoulli trials. In particular, we like to find a prediction interval $[L(X;n,m,\alpha),U(X;m,n,\alpha)]$ so that

$$P_{X,Y}(L(X;n,m,\alpha)\le Y\le U(X;m,n,\alpha))\ge 1-2\alpha.$$

* Corresponding author. Tel.: +1 337 482 5283; fax: +1 337 482 5346.
E-mail address: krishna@louisiana.edu (K. Krishnamoorthy).

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.11.021

As the conditional distribution of one of the variables given the sum $X+Y$ does not depend on the parameter p, prediction intervals can be constructed based on the conditional distribution. Thatcher (1964) noted that a PI for Y can be obtained from the conditional distribution of X given $X+Y=s$, which is hypergeometric (with sample size s, number of defects n, and the lot size $n+m$) and does not depend on p. Thatcher's method is similar in nature to that of Clopper and Pearson's (1934) fiducial approach for constructing exact confidence intervals for a binomial success probability. Knüsel (1994) has pointed out that the acceptance regions of two-sample tests (one-sided tests for comparing success probabilities of X and Y) are exact PIs for Y. We found that the PIs for the binomial case given in Knüsel (1994), and the one described in Aitchison and Dunsmore (1980, Section 5.5), are indeed the same as the exact PIs that can be obtained by Thatcher's (1964) method. Faulkenberry (1973) provided a general method of constructing a PI on the basis of the conditional distribution given a complete sufficient statistic. He has illustrated his general approach for constructing PIs for the exponential and Poisson distributions based on the conditional distribution of Y (the variable to be predicted) given the sum $X+Y$. Bain and Patel (1993) have applied Faulkenberry's approach to construct PIs for the binomial, negative binomial and hypergeometric distributions. Dunsmore (1976) has shown that Faulkenberry's approach is not so quite general as claimed, and provided a counter example. Knüsel (1994, Section 3) noted that Faulkenberry's approach is correct for the continuous case, but it is not exact for the Poisson case.

Apart from the conditional methods given in the preceding paragraph, there is an asymptotic approach proposed in Nelson (1982), which is also reviewed in Hahn and Meeker (1991). Nelson's PI is based on the asymptotic normality of a standard pivotal statistic, and it is easier to compute than the ones based on the conditional approaches. However, Wang's (2008) coverage studies and our studies (see Figs. 1b and 3) indicate that Nelson's PIs have poor coverage probabilities even for large samples. Wang considered a modified Nelson's PI in which a factor has to be determined so that the minimum coverage probability is close to the nominal confidence level. Wang has provided a numerical approach to find the factor. A reviewer has brought to our attention that Wang (2010) has proposed another closed-form PI for a binomial distribution. It should be noted that the exact and conditional PIs were not included in Wang's (2008, 2010) coverage studies.

All the methods for the binomial case are also applicable for the Poisson case as shown in the sequel.

In this article, we propose a closed form approximate PIs based on the "joint sampling approach" which is similar to the one used to find confidence interval in a calibration problem (e.g., see Brown, 1982, Section 1.2). The proposed PIs are simple to compute, and they are comparable to or better than the available PIs in most cases. Furthermore, we show that the conditional PIs are narrower than the corresponding exact PIs except for the extreme cases where the conditional PIs are not defined. We also show that the new PI is included in or identical to Wang's (2010) PI.

The rest of the article is organized as follows. In the following section, we review the conditional methods, Nelson's PIs, Wang's (2010) PIs, and present the new PIs based on the joint sampling approach for the binomial case. The exact coverage probabilities and expected widths of the PIs are evaluated. The results for the binomial case are extended to the Poisson distribution in Section 3. Construction of the binomial and Poisson PIs are illustrated using two examples in Section 4. Some concluding remarks are given in Section 5.

2. Binomial distribution

Let $X \sim$ binomial(n,p) independently of $Y \sim$ binomial(m,p). The problem is to find a $1-2\alpha$ prediction interval for Y based on X. The conditional distribution of X given the sum $X+Y=s$ is hypergeometric with the sample size s, number of "non-defects" n, and the lot size $n+m$. The conditional probability mass function is given by

$$P(X=x|X+Y=s,n,n+m)=\frac{\binom{n}{x}\binom{m}{y}}{\binom{m+n}{s}},\quad \max\{0,s-m\}\le x\le \min\{n,s\}.$$

Let us denote the cumulative distribution function (cdf) of X given $X+Y=s$ by $H(t;s,n,n+m)$. That is,

$$H(t;s,n,n+m)=P(X\le t|s,n,n+m)=\sum_{i=0}^{t}\frac{\binom{n}{i}\binom{m}{s-i}}{\binom{m+n}{s}}. \tag{1}$$

Note that the conditional cdf of Y given $X+Y=s$ is given by $H(t;s,m,n+m)$.

2.1. Binomial prediction intervals

2.1.1. The exact prediction interval

Thatcher (1964) developed the following exact PI on the basis of the conditional distribution of X given $X+Y$. Let x be an observed value of X. The $1-\alpha$ lower prediction limit L is the smallest integer for which

$$P(X\ge x|x+L,n,n+m)=1-H(x-1;x+L,n,n+m)>\alpha. \tag{2}$$

The $1-\alpha$ upper prediction limit U is the largest integer for which

$$H(x;x+U,n,n+m)>\alpha. \tag{3}$$

Furthermore, $[L, U]$ is a $1-2\alpha$ two-sided PI for Y. Thatcher (1964) has noted that, for a fixed (x,n,m), the probability (3) is a decreasing function of U, and so a backward search, starting from m, can be used to find the largest integer U for which the probability in (3) is just greater α. Similarly, we see that the probability in (2) is an increasing function of L, and so a forward search method, starting from a small value, can be used to find the smallest integer L for which this probability is just greater than α.

The exact PIs for extreme values of X are defined as follows. When $X=0$, the lower prediction limit for Y is 0, and the upper one is determined by (3); when $x=n$, the upper prediction limit is m, and the lower prediction limit is determined by (2).

2.1.2. The conditional predication interval

Bain and Patel (1993) used Faulkenberry's (1973) conditional approach to develop a PI for Y. As noted earlier, this PI is based on the conditional distribution of Y, conditionally given $X+Y$, and is described as follows. Recall that the conditional cdf of Y given $X+Y=s$ is given by $P(Y\le t|X+Y=s)=H(t;s,m,n+m)$. Let x be an observed value of X. The $1-\alpha$ lower prediction limit L for Y is the largest integer so that

$$H(L-1;x+L,m,n+m)\le\alpha, \tag{4}$$

and the $1-\alpha$ upper prediction limit U is the smallest integer for which

$$H(U;x+U,m,n+m)\ge 1-\alpha. \tag{5}$$

Furthermore, $[L, U]$ is the $1-2\alpha$ two-sided PI for Y. As noted by Dunsmore (1976), Faulkenberry's method, in general, is not exact. We also note that the above prediction interval is not defined at the extreme cases $X=0$ and $X=n$.

2.1.3. The Nelson prediction interval

Let $\widehat{p}=X/n$, $\widehat{q}=1-\widehat{p}$ and $\widehat{Y}=m\widehat{p}$. Using the variance estimate $\widehat{\mathrm{var}}(\widehat{Y}-Y)=m\widehat{p}\widehat{q}(1+m/n)$, Nelson (1982) proposed an approximate PI based on the asymptotic result that

$$\frac{\widehat{Y}-Y}{\sqrt{\widehat{\mathrm{var}}(\widehat{Y}-Y)}}=\frac{\widehat{Y}-Y}{\sqrt{m\widehat{p}\widehat{q}(1+m/n)}}\sim \mathrm{N}(0,1). \tag{6}$$

The resulting $1-2\alpha$ PI is given by $\widehat{Y}\pm z_{1-\alpha}\sqrt{m\widehat{p}\widehat{q}(1+m/n)}$, where z_α is the 100α percentile of the standard normal distribution. Notice that the above PI is not defined when X=0 or n. As X assumes these values with positive probabilities, the coverage probabilities of the above PI are expected to be much smaller than the nominal level when p is at the boundary (Wang, 2008). To overcome the poor coverage probabilities at the boundary, we can define the sample proportion as $\widehat{p}=0.5/n$ if X=0 and $(n-0.5)/n$ if X=n. This type of adjustments is commonly used to handle extreme cases while estimating the odds ratio or the relative risk involving two binomial distributions (e.g., see Agresti, 1999). Since we are predicting a discrete random variable, the PI is formed by the set of integer values of Y satisfying $|\widehat{Y}-Y|/\sqrt{m\widehat{p}\widehat{q}(1+m/n)}|<z_{1-\alpha}$, and is given by

$$\left[\left\lceil \widehat{Y}-z_{1-\alpha}\sqrt{\widehat{Y}(m-\widehat{Y})\left(\frac{1}{m}+\frac{1}{n}\right)}\right\rceil,\quad \left\lfloor \widehat{Y}+z_{1-\alpha}\sqrt{\widehat{Y}(m-\widehat{Y})\left(\frac{1}{m}+\frac{1}{n}\right)}\right\rfloor\right], \tag{7}$$

where $\lceil x\rceil$ is the smallest integer greater than or equal to x, and $\lfloor x\rfloor$ is the largest integer less than or equal to x.

2.1.4. Wang's prediction interval

Wang (2010) proposed a PI by employing an approach similar to the construction of Wilson's score CI for the binomial proportion p. Wang's PI is based on the result that

$$W(m,n,X,Y)=\frac{Y-m\widehat{p}}{\left[\left(\frac{X+Y+c/2}{m+n+c}\right)\left(1-\frac{X+Y+c/2}{m+n+c}\right)\frac{m(m+n)}{n}\right]^{1/2}}\sim \mathrm{N}(0,1), \tag{8}$$

approximately, where $c=z_{1-\alpha}^2$. The $1-2\alpha$ PI is the set of integer values of Y that satisfy $|W(m,n,X,Y)|<z_{1-\alpha}^2$. To write the PI explicitly, let

$$A=mn[2Xz_{1-\alpha}^2(n+m+z_{1-\alpha}^2)+(2X+z_{1-\alpha}^2)(m+n)^2],$$
$$B=mn(m+n)z_{1-\alpha}^2(m+n+z_{1-\alpha}^2)^2\{2(n-X)[n^2(2X+z_{1-\alpha}^2)+4mnX+2m^2X]+nz_{1-\alpha}^2[n(2X+z_{1-\alpha}^2)+3mn+m^2]\},$$
$$C=2n[(n+z_{1-\alpha}^2)(m^2+n(n+z_{1-\alpha}^2))+mn(2n+3z_{1-\alpha}^2)].$$

In terms of above notations, Wang's PI is given by $[\lceil L\rceil,\lfloor U\rfloor]$, where

$$(L,U)=\frac{A}{C}\pm\frac{\sqrt{B}}{C}. \tag{9}$$

2.1.5. The prediction interval based on the joint sampling approach

The PI that we propose is based on the complete sufficient statistic $X+Y$ for the binomial $(n+m,p)$ distribution. Define $\widehat{p}_{xy}=(X+Y)/(n+m)$, and $\widehat{\text{var}}(m\widehat{p}_{xy}-Y)=mn\widehat{p}_{xy}\widehat{q}_{xy}/(n+m)$, where $\widehat{q}_{xy}=1-\widehat{p}_{xy}$. We can find a PI for Y based on the asymptotic joint sampling distribution that

$$\frac{m\widehat{p}_{xy}-Y}{\sqrt{\widehat{\text{var}}(m\widehat{p}_{xy}-Y)}}=\frac{(mX-nY)}{\sqrt{mn\widehat{p}_{xy}\widehat{q}_{xy}(n+m)}}\sim \text{N}(0,1). \tag{10}$$

To handle the extreme cases of X, let us define X to be 0.5 when $X=0$, and to be $n-0.5$ when $X=n$. The $1-2\alpha$ prediction interval is the set of integer values of Y satisfying

$$\frac{(mX-nY)^2}{mn\widehat{p}_{xy}\widehat{q}_{xy}(n+m)}<z_{1-\alpha}^2. \tag{11}$$

Treating as if Y is continuous, we can find the roots of the quadratic equation $((mX-nY)^2)/mn\widehat{p}_{xy}\widehat{q}_{xy}(n+m)=z_{1-\alpha}^2$ as

$$\frac{\left[\widehat{Y}\left(1-\frac{z_{1-\alpha}^2}{m+n}\right)+\frac{mz_{1-\alpha}^2}{2n}\right]\pm z_{1-\alpha}\sqrt{\widehat{Y}(m-\widehat{Y})\left(\frac{1}{m}+\frac{1}{n}\right)+\frac{m^2z_{1-\alpha}^2}{4n^2}}}{\left(1+\frac{mz_{1-\alpha}^2}{n(m+n)}\right)}=(L,U), \tag{12}$$

say, where $\widehat{Y}=mX/n$ for $X=1,\ldots,n-1$; it is $0.5m/n$ when $X=0$ and $(n-0.5)m/n$ when $X=n$. The $1-2\alpha$ PI, based on the roots in (12), is given by $[\lceil L\rceil,\lfloor U\rfloor]$. We refer to this PI as the "joint sampling–prediction interval" (JS–PI).

2.2. Comparison studies and expected widths of binomial prediction intervals

The conditional PI is either included in or identical to the corresponding exact PI except for the extreme cases where the former PI is not defined (see Appendix A for a proof). In order to study the coverage probabilities and expected widths of the conditional PI, we take it to be the exact PI when X=0 or n. Specifically, the conditional PI is determined by (4) and (5) when $X=1,\ldots,n-1$, and by (2) or (3) when $X=n$ or 0. Between our JS–PI and Wang's (2010) PI, the former is included in or identical to Wang's (2010) PI (see Appendix B for a proof). There are several cases where our JS–PI is strictly shorter than the Wang PI as shown in the following table.

(n, m)	X	95% PIs	
		Wang's PI	JS–PI
(40, 20)	16	[3, 13]	[4, 13]
	33	[12, 20]	[12, 19]
(100, 50)	10	[1, 11]	[1, 10]
(300, 100)	18	[1, 12]	[2, 11]
(1000, 300)	278	[67,101]	[67, 100]
(3000, 100)	212	[2, 12]	[3, 12]

Note that for all the cases considered in the above table, our JS–PI is shorter than Wang's PI by only one unit.

To compare PIs on the basis of coverage probabilities, we note that, for a given (n,m,p,α), the exact coverage probability of a PI $[L(x,n,m,\alpha),U(x,n,m,\alpha)]$ can be evaluated using the expression

$$\sum_{x=0}^{n}\sum_{y=0}^{m}\binom{n}{x}\binom{m}{y}p^{x+y}(1-p)^{m+n-(x+y)}I_{[L(x,n,m,\alpha),U(x,n,m,\alpha)]}(y), \tag{13}$$

where $I_A(.)$ is the indicator function. For a good PI, its coverage probabilities should be close to the nominal level.

We evaluated the coverage probabilities of all five PIs using (13) for some sample sizes and confidence coefficient 0.95. In order to display the plots clearly, we plotted the coverage probabilities of Wang's (2010) PIs in (9) and those of the JS–PIs (12) in Fig. 1a. Even though we knew that the JS–PI is contained in Wang's PI, the coverage probabilities of these two PIs should be evaluated because both are approximate. It is clear from these plots that both PIs could be liberal, but their coverage probabilities are barely below the nominal level 0.95. As expected, for all sample size configurations considered, the coverage probabilities of Wang's PIs are greater than or equal to those of the JS–PIs. Thus, Wang's PIs are more conservative than the JS–PIs for some parameter space, especially when p is away from 0.5. As a result, Wang's PIs are expected to be wider than the JS–PIs. In view of these comparison results, we will not include Wang's PIs for further comparison study in the sequel.

The coverage probabilities of the PIs based on the exact, conditional, Nelson's PIs and joint sampling methods are plotted in Fig. 1b. We first observe from these plots that all the PIs, except the Nelson PI, are conservative when p is close to zero or one. The exact PIs are too conservative even for large samples; see the coverage plots for n=m=50, n=50, m=20 and n=100, m=20.

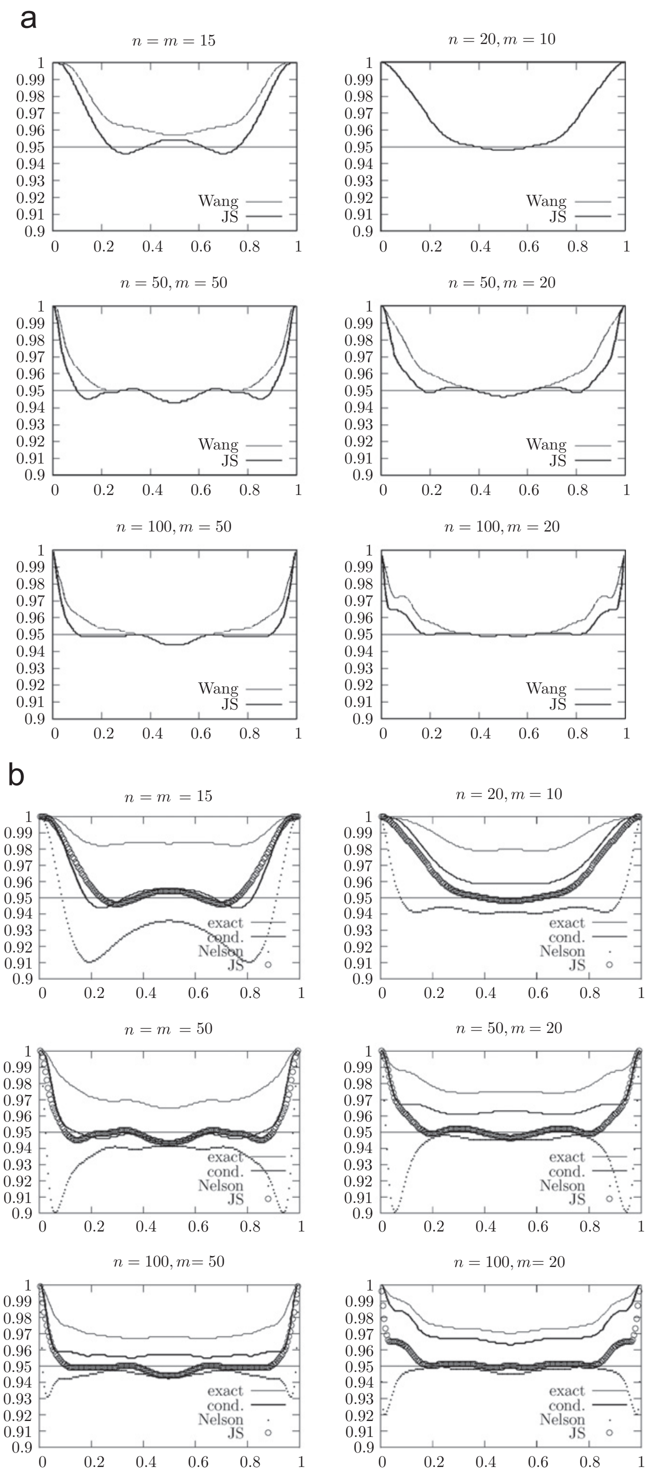

Fig. 1. Coverage probabilities of 95% binomial PIs as a function of p.

As expected, the conditional PIs are less conservative than the exact PIs for all the cases considered. Furthermore, the coverage probabilities of the conditional PIs are seldom fall below the nominal level 0.95; see the plots for $n=m=15$ and $n=m=50$. Nelson's PIs are in general liberal, especially when the values of the parameter are at the boundary. Overall, the exact PIs are too conservative, and Nelson's PIs are liberal even for large samples. The conditional PI and the new one based on the joint sampling approach (JS–PI) are

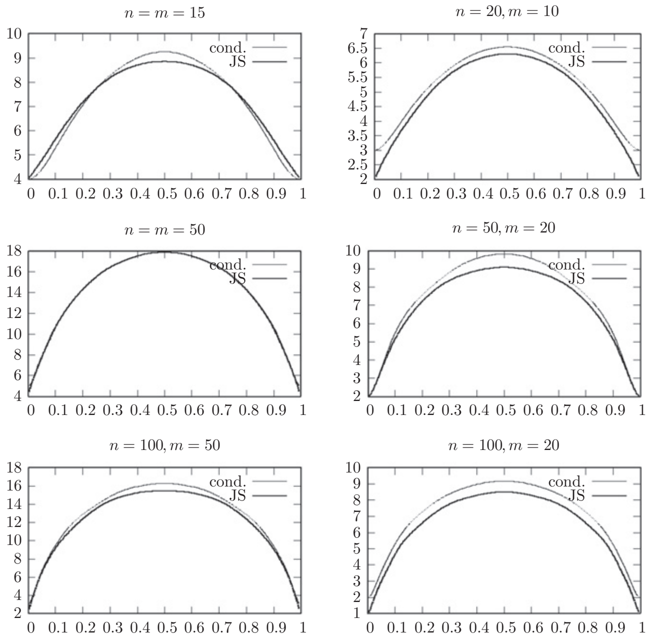

Fig. 2. Expected widths of 95% binomial prediction intervals as a function of p.

comparable in some situations (see the plots for n=m=15 and n=m=50), and the latter exhibits better performance than the former in other cases. In particular, in the most common case where the past and current samples are larger than the future sample, the new PIs are better than the corresponding conditional PIs (see the plots for the cases n=50, m=20 and n=100, m=20).

As the conditional PI and the new PI are the ones that control the coverage probabilities satisfactorily, we compare only these PIs with respect to expected widths. We evaluate the expected widths of these PIs for the sample size configurations for which coverage probabilities were computed earlier. The plots of the expected widths are given in Fig. 2. These plots indicate that the expected widths of the new PIs are comparable to or smaller than those of the conditional PIs for all the cases considered. In general, the new prediction interval is preferable to the conditional PI with respect to coverage properties and expected widths.

3. Poisson distribution

Let X be the total counts in a sample of size n from a Poisson distribution with mean λ. Note that $X\sim$ Poisson $(n\lambda)$. Let Y denote the future total counts that can be observed in a sample of size m from the same Poisson distribution so that $Y\sim$ Poisson $(m\lambda)$. Note that the conditional distribution of X, conditionally given $X+Y=s$, is binomial with number of trials s and the success probability $n/(n+m)$, binomial$(s,n/(m+n))$, and the conditional distribution of Y given the sum X+Y is binomial$(s,m/(m+n))$. Let us denote the cumulative distribution function of a binomial random variable with the number of trials N and success probability π by $B(x;N,\pi)$.

3.1. The exact prediction interval

As noted earlier, the exact PI is based on the conditional distribution of X given X+Y. Let x be an observed value of X. The smallest integer L that satisfies

$$1-B(x-1;x+L,n/(n+m))>\alpha \tag{14}$$

is the $1-\alpha$ lower prediction limit for Y. The $1-\alpha$ upper prediction limit U is the largest integer for which

$$B(x;x+U,n/(n+m))>\alpha. \tag{15}$$

For $X=0$, the lower prediction limit is defined to be zero, and the upper prediction limit is determined by (15).

3.2. The conditional prediction interval

The conditional approach is based on the conditional distribution of Y given the sum X+Y. The $1-\alpha$ lower prediction limit is the largest integer L for which

$$B(L-1;x+L,m/(m+n))\leq\alpha, \tag{16}$$

where x is an observed value of X. The $1-\alpha$ upper prediction limit U is the smallest integer for which

$$B(U; x+U, m/(m+n)) \geq 1-\alpha. \tag{17}$$

3.3. The Nelson prediction interval

Let $\widehat{\lambda} = X/n$, and $\widehat{\mathrm{var}}(m\widehat{\lambda}-Y) = m^2\widehat{\lambda}(1/n+1/m)$. Nelson's approximate PI is based on the asymptotic result that $(m\widehat{\lambda}-Y)/\sqrt{\widehat{\mathrm{var}}(m\widehat{\lambda}-Y)} \sim N(0,1)$, and is given by

$$[\lceil L \rceil, \lfloor U \rfloor] \quad \text{with } [L,U] = \widehat{Y} \pm z_{1-\alpha}\sqrt{m\widehat{Y}\left(\frac{1}{m}+\frac{1}{n}\right)}, \tag{18}$$

where $\widehat{Y} = mX/n$, for $X = 1,2,\ldots$, and is $0.5m/n$ when $X=0$.

3.4. The prediction interval based on the joint sampling approach

Let $\widehat{\lambda}_{xy} = (X+Y)/(m+n)$. As in the binomial case, using the variance estimate $\widehat{\mathrm{var}}(m\widehat{\lambda}_{xy}-Y) = mn\widehat{\lambda}_{xy}/(m+n)$, we consider the quantity

$$\frac{m\widehat{\lambda}_{xy}-Y}{\sqrt{\widehat{\mathrm{var}}(m\widehat{\lambda}_{xy}-Y)}} = \frac{(mX-nY)}{\sqrt{mn(X+Y)}},$$

whose asymptotic joint distribution is the standard normal. In order to handle the zero count, we take X to be 0.5 when it is zero. The $1-2\alpha$ PI is determined by the roots (with respect to Y) of the quadratic equation $(m\widehat{\lambda}-Y)^2/[\widehat{\lambda}_{xy}m(1+m/n)] = z_{1-\alpha}^2$. Based on these roots, the $1-2\alpha$ PI is given by

$$[\lceil L \rceil, \lfloor U \rfloor] \quad \text{with } [L,U] = \widehat{Y} + \frac{mz_{1-\alpha}^2}{2n} \pm z_{1-\alpha}\sqrt{m\widehat{Y}\left(\frac{1}{m}+\frac{1}{n}\right) + \frac{m^2 z_{1-\alpha}^2}{4n^2}}, \tag{19}$$

where $\widehat{Y} = mX/n$, for $X = 1,2,\ldots$, and is $0.5m/n$ for X=0.

3.5. Coverage probabilities and expected widths of Poisson prediction intervals

As in the binomial case, it can be shown that the conditional PIs are either included in or identical to the corresponding exact PIs except for the case of X=0 where the former PIs are not defined. This can be proved along the lines for the binomial case given in Appendix A, and so the proof is omitted.

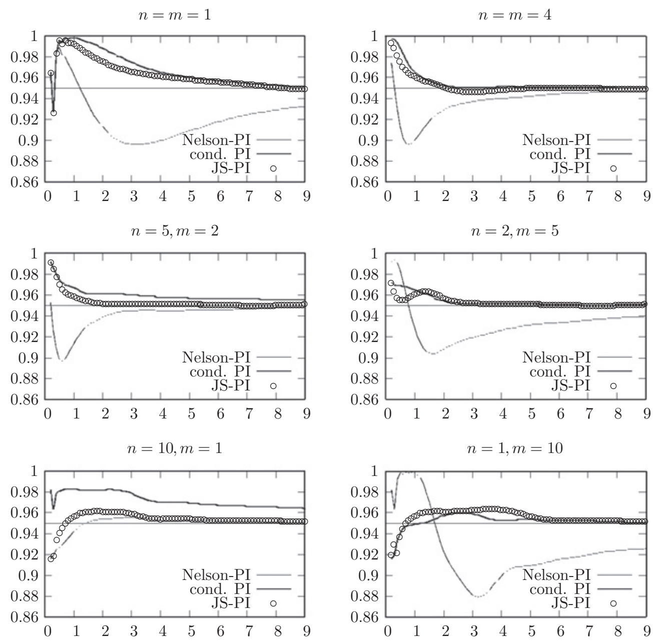

Fig. 3. Coverage probabilities of 95% prediction intervals for Poisson distributions as a function of λ.

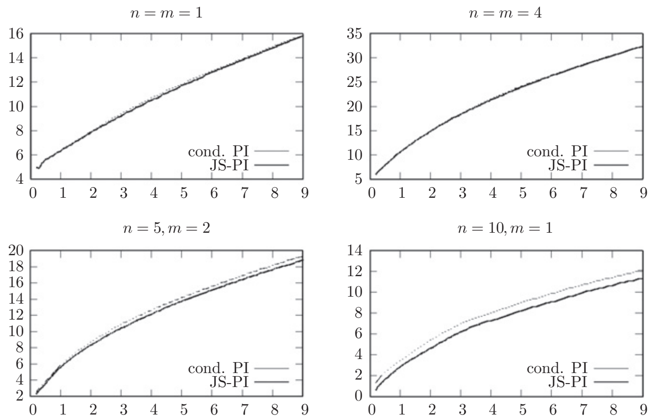

Fig. 4. Expected widths of 95% Poisson prediction intervals as a function of λ.

In order to study the unconditional coverage probabilities and expected widths of the conditional PIs, we take them to be exact PIs when X=0.

For a given (n,m,λ,α), the exact coverage probability of a Poisson PI $[L(x,n,m,\alpha),U(x,n,m,\alpha)]$ can be evaluated using the expression

$$\sum_{x=0}^{\infty}\sum_{y=0}^{\infty}\frac{e^{-\lambda(n+m)}(n\lambda)^x(m\lambda)^y}{x!y!}I_{[L(x,n,m,\alpha),U(x,n,m,\alpha)]}(y).$$

The coverage probabilities of all PIs, except the exact one, are plotted in Fig. 3 for some values of (n,m). The exact PI is not included in the plots because it exhibited similar performance (excessively conservative) as in the binomial case, and it is inferior to the conditional PI in terms of coverage probabilities and expected widths. Regarding other PIs, we see from these six plots in Fig. 3 that Nelson's PIs are often liberal. The conditional PI and the new JS–PI behave similarly when n=m, and the former is more conservative than the latter when $n>m$ (see the plots for (n=5, m=2) and (n=10, m=1)). Both PIs could be occasionally liberal, especially when λ is very small.

As the conditional PI and the new one based on the joint sampling approach are the ones satisfactorily control the coverage probabilities, we further compare them with respect to expected widths. The plots of expected widths are given in Fig. 4. These plots clearly indicate that the expected widths of these two PIs are not appreciably different for the situations they have similar coverage probabilities. On the other hand, for the cases where the conditional PIs are conservative, the new PIs have shorter expected widths; see the plots of ($n=5,m=2$) and (n=10, m=1).

We also evaluated coverage probabilities and expected widths of PIs with confidence coefficients 0.90 and 0.99. The results of comparison are similar to those of 95% PIs, and so they are not reported here. On an overall basis, the new PI is preferable to other PIs in terms of coverage probabilities and precisions.

4. Examples

4.1. An example for binomial prediction intervals

This example is adapted from Bain and Patel (1993). Suppose for a random sample of n=400 devices tested, x=20 devices are unacceptable. A 90% prediction interval is desired for the number of unacceptable devices in a future sample of m=100 such devices. The sample proportion of unacceptable devices is $\widehat{p}=20/400=0.05$ and $\widehat{Y}=m\times\widehat{p}=5$. The normal critical value required to compute Nelson's PI and the new PI is $z_{0.95}=1.645$. Nelson's PI given in (7) is [1,9]. The new JS–PI in (12) and Wang's PI in (9) are [2,9]. To compute the conditional PI, we can use the new prediction limits just obtained as initial guess values. In particular, using L=2 as the initial guess value, we compute the probability in (4) as $H(L-1=1,x+L=22,100,500)=0.0446$ at L=2; $H(L-1=2,x+L=23,100,500)=0.1272>0.05$ at L=3. So the lower endpoint is 2. Similarly, we can find the upper endpoint as 9. Thus, the conditional PI is also [2,9]. To compute the exact PI, we note that the probability in (2) at L=2 is $H(20;22,400,500)=0.1485$, and at L=1 is $1-H(19;21,400,500)=0.0540$. Thus, the lower endpoint of the exact PI is 1. The upper endpoint can be similarly obtained as 10. Thus, exact PI is [1,10], which is the widest among all PIs.

4.2. An example for Poisson prediction intervals

We shall use the example given in Hahn and Chandra (1981) for illustrating the construction of PIs for a Poisson distribution. The number of unscheduled shutdowns per year in a large population of systems follows a Poisson distribution with mean λ. Suppose there were 24 shutdowns over a period of 5 years, and we like to find 90% PIs for the number of unscheduled shutdowns over the next year. Here n=5, x=24 and m=1. The maximum likelihood estimate of λ is given by $\widehat{\lambda} = \frac{24}{5} = 4.8$.

Nelson's PI in (7) is computed as $[\lceil 4.8-3.95\rceil, \lfloor 4.8+3.95\rfloor] = [1,8]$. The new PI based on the joint sampling approach given in (19) is [2, 9]. To compute the conditional PI, we note that $m/(m+n) = 0.1667$ and

$$B(8;24+8,0.1667) = 0.9271 \quad \text{and} \quad B(9;24+9,0.1667) = 0.9621.$$

So 9 is the upper endpoint of the 90% conditional PI. Furthermore, $B(a-1;24+a,0.1667) = 0.0541$ at $a = 2$ and $B(a-1;24+a,0.1667) = 0.0105$ at $a = 1$, and so the lower endpoint of the 90% PI is 1. Thus, the conditional PI is [1, 9]. To find the lower endpoint of the exact PI using (14), we note that $n/(n+m) = 5/6 = 0.8333$, and $1-B(23;24+1,0.8333) = 0.0628$. It can be verified that L=1 is the smallest integer for which $1-B(23;24+L,0.8333 > 0.05$. Similarly, using (15), we can find the upper endpoint of the 90% exact PI as 9. Thus, the exact PI and the conditional PI are the same [1,9] whereas the new PI is [2,9].

5. Concluding remarks

It is now well-known that the classical exact methods for discrete distributions are often too conservative producing confidence intervals that are unnecessarily wide or tests that are less powerful. Examples include Clopper and Pearson (1934) confidence interval for a binomial proportion, Fisher's exact test for comparing two proportions, Przyborowski and Wilenski's (1940) conditional test for the ratio of two Poisson means, and Thomas and Gart's (1977) method for estimating the odds ratio involving two binomial distributions; see Agresti and Coull (1988), Agresti (1999) and Brown et al. (2001). Many authors have recommended alternative approximate approaches for constructing confidence intervals with satisfactory coverage probabilities and good precision. In this article, we have established similar results for constructing PIs for a binomial or Poisson distribution. Furthermore, we showed that the conditional PIs are narrower than the exact PIs, and so the former PIs have shorter expected width. The joint sampling approach produces PIs for both binomial and Poisson distributions with good coverage properties and precisions. In terms of simplicity and accuracy, the new PIs are preferable to others, and can be recommended for applications.

Acknowledgement

The authors are grateful to a reviewer for providing useful comments and suggestions.

Appendix A

We shall now show that, for the binomial case, the conditional PI is always included in or equal to the exact PI. Let L_c and L_e be the lower endpoints of the $1-2\alpha$ conditional PI and exact PI, respectively. To show that L_c is always greater than or equal to L_e, we recall that L_c is the largest integer so that $P(Y \le L_c-1|x+L_c,m,n+m) \le \alpha$. This is the probability of observing L_c-1 or fewer defects in a sample of size $x+L_c$ from a lot of size $n+m$ that contains m defects. This is the same as the probability of observing at least x+1 non-defects in a sample of size $x+L_c$ from a lot of size n+m that contains n non-defects. That is, $P(Y \le L_c-1|x+L_c,m,n+m) = P(X \ge x+1|x+L_c,n,n+m) \le \alpha$. Note that x must be the smallest integer for which the above inequality holds. Therefore,

$$P(X \ge x+1|x+L_c,n,n+m) \le \alpha \;\Rightarrow\; P(X \ge x|x+L_c,n,n+m) > \alpha.$$

Thus, L_c satisfies the condition in (2), and so it is greater than or equal to L_e. To show that $L_c > L_e$ for some sample sizes, recall that L_e is the smallest integer for which $P(X \ge x|x+L_e,n,n+m) > \alpha$. Arguing as before, it is easy to see that $P(X \ge x|x+L_e,n,n+m) = P(Y \le L_e|x+L_e,m,n+m) > \alpha$. Since L_e is the smallest integer for which this inequality holds, $P(Y \le L_e-1|x+L_e-1,m,n+m) \le \alpha$. As noted in Thatcher (1964), the hypergeometric cdf is a decreasing function of the sample size while other quantities are fixed, and so

$$P(Y \le L_e-1|x+L_e,m,n+m) < \alpha.$$

However, L_e is not necessarily the largest integer satisfying the above inequality, and so it could be smaller than L_c. In fact, there are many situations where $L_c > L_e$. For examples, when n=40, m=32 and x=22, the 95% exact and conditional PIs intervals are [10,25] and [11,24], respectively; when $(n,m,x) = (36,22,18)$, they are [5,17] and [6,16], respectively.

We shall now show that the upper endpoint U_c of a $1-2\alpha$ conditional PI is included in the corresponding exact PI. Recall that U_c is the smallest integer for which

$$P(Y \le U_c|x+U_c,m,n+m) \ge 1-\alpha. \tag{20}$$

As argued earlier, the above probability is the same as the probability of observing at least x non-defects in a sample of size x+U_c when the number of non-defects is n. That is, $P(Y \le U_c|x+U_c,m,n+m) = P(X \ge x|x+U_c,n,n+m)$. Because U_c is the smallest integer satisfying (20), x must be the largest integer satisfying $P(X \ge x|x+U_c,n,n+m) \ge 1-\alpha$. Thus, if U_c satisfies (20), then

$$P(X \geq x|x+U_c,n,n+m) \geq 1-\alpha \Leftrightarrow P(X \leq x-1|x+U_c,n,n+m) \leq \alpha, \tag{21}$$

which implies that $P(X \leq x|x+U_c,n,n+m) > \alpha$. In other words, U_c satisfies (3), and so it is included in the exact PI. Furthermore, using the argument similar to the one given for the lower endpoint, it can be shown that there are situations where U_c could be strictly less than the upper endpoint of the corresponding exact PI.

Appendix B

We here show that the JS–PI in (12) is included in or identical to Wang's PI determined by (8). For a given (n,m,X), let Y_0 be a value that is in the JS–PI. To show that Y_0 is included in Wang's PI, it is enough to show that if Y_0 satisfies (12) then it also satisfies $|W(n,m,X,Y_0)| < z_{1-\alpha}$, where $W(n,m,X,Y)$ is defined in (8). Equivalently, we need to show that

$$\frac{X+Y_0}{m+n}\left(1-\frac{X+Y_0}{m+n}\right) \leq \frac{X+Y_0+c/2}{m+n+c}\left(1-\frac{X+Y_0+c/2}{m+n+c}\right),$$

where $c = z_{1-\alpha}^2$. Let $d = X+Y_0$ and $N = m+n$. It is easy to check that the above inequality holds if and only if

$$d^2 - dN + \frac{N^2}{4} \geq 0.$$

The above quadratic function of d attains its minimum at $d = N/2$, and the minimum value is zero. Thus, the JS–PI is included in or identical to Wang's PI.

References

Aitchison, J., Dunsmore, I.R., 1980. Statistical Prediction Analysis. Cambridge University Press, Cambridge, UK.
Agresti, A., 1999. On logit confidence intervals for the odds ratio with small samples. Biometrics 55, 597–602.
Agresti, A., Coull, B.A., 1988. Approximate is better than "exact" for interval estimation of binomial proportion. The American Statistician 52, 119–125.
Bain, L.J., Patel, J.K., 1993. Prediction intervals based on partial observations for some discrete distributions. IEEE Transactions on Reliability 42, 459–463.
Brown, P.J., 1982. Multivariate calibration. Journal of the Royal Statistical Society, Series B 44, 287–321.
Brown, L.D., Cai, T., Das Gupta, A., 2001. Interval estimation for a binomial proportion (with discussion). Statistical Science 16, 101–133.
Clopper, C.J., Pearson, E.S., 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26, 404–413.
Davis, C.B., McNichols, R.J., 1987. One-sided intervals for at least p of m observations from a normal population on each of r future occasions. Technometrics 29, 359–370.
Dunsmore, I.R., 1976. A note on Faulkenberry's method of obtaining prediction intervals. Journal of the American Statistical Association 71, 193–194.
Faulkenberry, G.D., 1973. A method of obtaining prediction intervals. Journal of the American Statistical Association 68, 433–435.
Gibbons, R.D., 1987. Statistical prediction intervals for the evaluation of ground-water quality. Ground Water 25, 265–455.
Hahn, G.J., Meeker, W.Q., 1991. Statistical Intervals: A Guide to Practitioners. Wiley, New York.
Hahn, G.J., Chandra, R., 1981. Tolerance intervals for Poisson and binomial variables. Journal of Quality Technology 13, 100–110.
Knüsel, L., 1994. The prediction problem as the dual form of the two-sample problem with applications to the Poisson and the binomial distribution. The American Statistician 48, 214–219.
Nelson, W., 1982. Applied Life Data Analysis. Wiley, New York.
Przyborowski, J., Wilenski, H., 1940. Homogeneity of results in testing samples from Poisson series. Biometrika 31, 313–323.
Thatcher, A.R., 1964. Relationships between bayesian and confidence limits for prediction. Journal of the Royal Statistical Society, Series B 26, 176–192.
Thomas, D.G., Gart, J.J., 1977. A table of exact confidence limits for differences and ratios of two proportions and their odds ratios. Journal of the American Statistical Association 72, 73–76.
Wang, H., 2008. Coverage probability of prediction intervals for discrete random variables. Computational Statistics and Data Analysis 53, 17–26.
Wang, H., 2010. Closed form prediction intervals applied for disease counts. The American Statistician 64, 250–256.