18.443 Statistics for Applications
Spring 2009

CONFIDENCE INTERVALS FOR THE BINOMIAL PARAMETER p
18.443, R. M. Dudley

1. Introduction

First, here is some notation for binomial probabilities. Let X be the number of successes in n independent trials with probability p of success on each trial. Let $q \equiv 1-p$. Then we know that $EX = np$, the variance of X is npq where $q = 1-p$, and so the basic variance when $n = 1$ (Bernoulli distribution) is pq. For $k = 0, 1, ..., n$,

$$P(X = k) \ = \ b(k, n, p) \ := \ \binom{n}{k} p^k q^{n-k}$$

where $:=$ means "equals by definition." Let

$$B(k, n, p) \ := \ P(X \leq k) \ = \ \sum_{j=0}^{k} b(j, n, p),$$

$$E(k, n, p) \ := \ P(X \geq k) \ = \ \sum_{j=k}^{n} b(j, n, p).$$

$B(k, n, p)$ can be evaluated in R as pbinom(k, n, p). Thus $E(k, n, p)$ would be: $1-$ pbinom$(k-1, n, p)$.

To estimate a real parameter θ, such as μ, σ, or σ^2 for a normal $N(\mu, \sigma^2)$ distribution, or for the binomial distribution (for each $n \geq 1$) $\theta = p$, or the Poisson $\theta = \lambda$, an *interval estimator* is a pair of real-valued statistics $a(X)$ and $b(X)$ such that $a(X) \leq b(X)$ for all possible observations X. Let P_θ denote probability when θ is the true value of the parameter.

1.1. **Coverage probabilities.** The *coverage probability* for a given interval estimator $[a(\cdot), b(\cdot)]$ and θ is defined as

$$\kappa(\theta, a(\cdot), b(\cdot)) \ = \ P_\theta[a(X) \leq \theta \leq b(X)]. \qquad (1)$$

For $0 < \alpha < 1/2$ (usually $\alpha \leq 0.1$), I will say here that an interval estimator $[a(\cdot), b(\cdot)]$ is a *precise* $1-\alpha$ or $100(1-\alpha)\%$ *confidence interval* for θ if the coverage probability exactly equals $1-\alpha$ for all θ. As we've seen, for normal distributions $N(\mu, \sigma^2)$, and $n \geq 2$ i.i.d. observations, there are precise confidence intervals for the variance σ^2 based on the χ^2 distribution, and also precise confidence intervals for the mean, based on the t distribution. But for the binomial case there are no precise confidence intervals. The binomial random variable X has just $n+1$ possible values $0, 1, ..., n$, so for any interval estimator, there are just $n+1$ possible values $a(j)$ of the left endpoint and $b(j)$ of the right endpoint for $j = 0, 1, ..., n$. The coverage probability will take a jump upward as p crosses from below to above each $a(j)$ and downward as it crosses each $b(j)$ (unless possibly some $a(i)$ and $b(j)$ coincide). So the coverage probability in general is not constant and is not even a continuous function of p.

I'll say that an interval estimator $[a(\cdot), b(\cdot)]$ is a *secure* $1-\alpha$ or $100(1-\alpha)\%$ *confidence interval* for θ if the coverage probability is always at least $1-\alpha$,

$$\kappa(\theta, a(\cdot), b(\cdot)) \;\geq\; 1-\alpha \tag{2}$$

for all possible θ.

Comparing with terminology often used in the literature, the word "exact" has often been used for what I here call "secure," whereas "conservative" has been used to mean at least secure, or with a qualifier such as "overly conservative," as will be said sometimes herein, to indicate that the coverage probabilities are larger than $1-\alpha$ by an excessively or unnecessarily large amount.

On the other hand, the authors of beginning statistical texts mainly agree on what a confidence interval is if it is precise (as defined above), but if such an interval does not exist, most seem to have in mind a relatively vague notion of an interval estimator whose coverage probabilities are as close as practicable to $1-\alpha$ for as wide a range of θ as practicable. The texts' authors evidently want the endpoints $a(\cdot)$ and $b(\cdot)$ of the intervals to be relatively easy to compute.

1.2. **The plug-in interval.** By far the most popular interval for the binomial p (in beginning textbooks, not necessarily among mathematical statisticians) is the one defined as follows. Let $\hat{p} := X/n$ and $\hat{q} := 1-\hat{p}$. The *plug-in* interval estimator for p is defined by

$$[\hat{p} - z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}, \hat{p} + z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}], \tag{3}$$

recalling that z_β is defined so that if Z has a $N(0,1)$ distribution then $P(Z \geq z_\beta) = \beta$. A quick way to see a fault in the plug-in interval it is to see what happens when $X=0$, when $\hat{p}=0$, so $a(0)=0$ (which is fine) but also $b(0)=0$, which is very bad, because as $p>0$ decreases down toward 0, the coverage probability $\kappa(p)$ converges to 0. Symmetrically, if $X=n$ then $\hat{q}=0$ and $a(n)=b(n)=1$ so $\kappa(p)\to 0$ as $p \uparrow 1$.

1.3. **Clopper-Pearson intervals.** The *Clopper-Pearson* $100(1-\alpha)\%$ interval estimator for p is the interval $[a(X), b(X)] \equiv [a_{CP}(X), b_{CP}(X)]$ where $a_{CP}(X) = a_{CP}(X,n,\alpha)$ and $b_{CP}(X) = b_{CP}(X,n,\alpha)$ are such that if the true p were $a_{CP}(X)$, and $0 < X \leq n$, then if V has a binomial(n,p) distribution, the probability that $V \geq X$ would be $\alpha/2$, in other words,

$$E(X, n, a_{CP}(X)) \;=\; \alpha/2. \tag{4}$$

If $X=0$ then $a_{CP}(0) := 0$. Symmetrically, if the true p were $b_{CP}(X)$ and $0 \leq X < n$, and U has a binomial(n,p) distribution, the probability that $U \leq X$ would be $\alpha/2$, in other words,

$$B(X, n, b_{CP}(X)) \;=\; \alpha/2, \tag{5}$$

and if $X=n$ then $b_{CP}(n) := 1$.

The Clopper-Pearson confidence interval for p if $0 < X < n$ is defined in a way very analogous to the way 2-sided precise confidence intervals are for the normal μ and σ^2.

This makes the Clopper-Pearson intervals intuitive, and they have been called "exact," but they are not precise.

Theorem 1. *The Clopper-Pearson intervals are secure (for $0 < \alpha < 1$), in fact their coverage probabilities $\kappa(p) > 1-\alpha$ for all p. Moreover, for each n and α, $\inf_{0\le p\le 1} \kappa(p) > 1-\alpha$, i.e. for some $\delta = \delta(n,\alpha) > 0$, $\kappa(p) \ge 1-\alpha+\delta$ for all p.*

Proof. To see that the intervals are monotone, i.e. $a_{CP}(0) = 0 < a_{CP}(1) < \cdots < a_{CP}(n) < 1$ and $0 < b_{CP}(0) < b_{CP}(1) < \cdots < b_{CP}(n) = 1$, consider for example that $E(j,n,a(j)) = \alpha/2 = E(j+1,n,a(j+1))$, so to produce equal probability, for a larger number of successes ($X \ge j+1$, vs. $X \ge j$), $a(j+1)$ must be larger than $a(j)$. The situation for the $b(j)$ is symmetrical.

For any p with $0 \le p \le 1$, let $J = J(p)$ be the smallest j such that $p \le b(j)$, and let $K = K(p)$ be the largest k such that $a(k) \le p$. The definition makes sense since $a_{CP}(0) = 0$ and $b_{CP}(n) = 1$. Then $a(K) := a_{CP}(K) \le p \le b(J) := b_{CP}(J)$. By monotonicity, $a(i) \le a(K) \le p$ for $i = 0,1,...,K$ and $b(i) \ge b(J)$ for $J \le i \le n$. Let X be binomial (n,p). Then from the definitions, $\Pr_p(X > K) = 0$ if $K = n$, otherwise it equals $E(K+1,n,p) < E(K+1,n,a(K+1)) = \alpha/2$ since $E(r,n,p)$ is increasing in p and $p < a(K+1)$. Symmetrically, $\Pr_p(X < J) < \alpha/2$. It follows (using $\alpha < 1$) that $J \le K$ so $p \in [a(r),b(r)]$ for all $r = J,...,K$ by monotonicity. If $r > K$ then $a(r) > p$ by definition of $K(p)$, while if $r < J$ then $b(r) < p$ by definition of $K(p)$. Thus $p \in [a(r),b(r)]$ if and only if $J \le r \le K$, and

$$\kappa(p) = \sum_{r=J}^{K} \Pr_p(X = r) = 1 - E(K+1),n,p) - B(J-1,n,p) > 1-\alpha/2-\alpha/2 = 1-\alpha,$$

where the first (second) $\alpha/2$ is replaced by 0 if $K = n$ or $J = 0$ respectively. Also, $E(k+1,n,p)$ can only approach $\alpha/2$, for the given values of $J(p)$ and $K(p)$, for $p \uparrow a(K+1)$, and $B(J-1,n,p)$ can only approach $\alpha/2$ as $p \downarrow b(J-1)$. These things cannot happen simultaneously, so $\kappa(p)$ cannot approach $1-\alpha$ and must be $\ge 1-\alpha+\delta$ for some $\delta > 0$. This finishes the proof. □

The Clopper-Pearson intervals are overly conservative in that, for example, for $0 \le p \le b(0)$, however $b(0)$ is defined, if the left endpoints $a(j) = a_{CP}(j)$, $\kappa(p) \ge 1-(\alpha/2)$. This is illustrated in Table 0 where for $n = 20$ and $\alpha = 0.01$, $b(0) = b_{CP}(0) \doteq 0.2327$, and for $p \le b(0)$, which holds for $p = a(j)$ for $j \le 10$, all coverage probabilities shown are $\ge 0.995 = 1-\alpha/2$.

One might easily be tempted, if one observes $X = 0$ and notes that the resulting interval will be one-sided in the sense that $a(0) = 0$, to choose $b(0)$ such that if $p = b(0)$, the probability of observing $X = 0$ would be α, rather than $\alpha/2$ as in definition (5). That can lose the secure property, however: for example if $n = 20$ and $\alpha = 0.01$, $b(0) \doteq 0.2327$ would be replaced by the smaller $b(0) \doteq 0.2057 < a_{CP}(10)$, and we would get $\kappa(b(0)+) \doteq 0.9868$ and $\kappa(a(10)-) \doteq 0.9876$, both less than $0.99 = 1-\alpha$. Likewise, for the secure property, if $X = n$, we need to keep $a(n)$ as it is by (4) rather than replace $\alpha/2$ by α.

The coverage probability $\kappa(p)$ can be close to $1-\alpha$ if the interval between $b(J-1)$ and $a(K+1)$, which contains p, is a short one, as seen in Table 0 for $b(2) < p < a(14)$ where $\kappa(p) \doteq 0.9904$ and so the $\delta(20, 0.01)$ as defined in Theorem 1 is about 0.0004.

Table 0. Clopper-Pearson confidence intervals for $n = 20$, $\alpha = 0.01$ *and their coverage probabilities.*

At each endpoint $p = a(j) = a_{CP}(j)$ or $b(j) = b_{CP}(j)$, the coverage probability has a left limit $\kappa(p-) = \lim_{r\uparrow p}\kappa(r)$ (except for $p = a(0) = 0$) and a right limit $\kappa(p+) = \lim_{r\downarrow p}\kappa(r)$ (except at $p = 1$, not shown). Actually $\kappa(p+) = \kappa(p)$ if $p = a(j)$ for some j and $\kappa(p-) = \kappa(p)$ if $p = b(j)$ for some j.

For $p < b(0)$, we have $J(p)$ as defined in the proof of Theorem 1 equal to 0, and thus the coverage probability as shown in that proof is $B(K(p), n, p) > 1 - \alpha/2 = 0.995$ in this case as seen in the table.

The table only covers endpoints $p < 0.5$, but the rest are determined by $a(n-j) = 1 - b(j)$ for $j = 0, 1, 2, 3$ and $b(n-k) = 1 - a(k)$ for $k = 0, 1, ..., 16$, and the coverage probabilities for $1/2 \leq p \leq 1$ would be symmetric to those shown for $p \leq 1/2$.

On an interval between two consecutive endpoints, the coverage probability will be $\kappa(p) = B(K, n, p) - B(J-1, n, p)$ as in the proof of Theorem 1. Differentiating this with respect to p, we get telescoping sums, and the derivative equals a positive function times a decreasing function of p, which is positive when p is small and negative as p approaches 1. Thus $\kappa(p)$ cannot have an internal relative minimum on such an interval, and its smallest values must be those on approaching an endpoint, which are shown in the following table.

	endpoint	$\kappa(p-)$	$\kappa(p+)$
$a(0)$	0.0000	—	1.0000
$a(1)$	0.00025	0.9950	1.0000
$a(2)$	0.0053	0.9950	0.9998
$a(3)$	0.0176	0.9950	0.9996
$a(4)$	0.0358	0.9950	0.9994
$a(5)$	0.0583	0.9950	0.9993
$a(6)$	0.0846	0.9950	0.9991
$a(7)$	0.1139	0.9950	0.9990
$a(8)$	0.1460	0.9950	0.9989
$a(9)$	0.1806	0.9950	0.9988
$a(10)$	0.2177	0.9950	0.9988
$b(0)$	0.2327	0.9979	0.9929
$a(11)$	0.2572	0.9924	0.9962
$a(12)$	0.2991	0.9942	0.9979
$b(1)$	0.3171	0.9973	0.9927
$a(13)$	0.3434	0.9925	0.9962
$b(2)$	0.3871	0.9947	0.9904
$a(14)$	0.3904	0.9904	0.9942
$a(15)$	0.4402	0.9938	0.9976
$b(3)$	0.4495	0.9976	0.9935
$a(16)$	0.4934	0.9934	0.9974

1.4. **Adjusted Clopper-Pearson intervals.** One can adjust the endpoints by, whenever $a'(k) := a_{CP}(k, n, 2\alpha) \leq b(0)$, which will occur for $k \leq k_0$ for some $k_0 = k_0(\alpha)$, replacing $a_{CP}(k)$ by the larger $a'(k)$. Symmetrically, for these k, $b_{CP}(n-k)$ is replaced by $1-a'(k)$. Then the intervals remain secure, but now as $p \uparrow a(k)$ for $1 \leq k \leq k_0$, $\kappa(p)$ will approach $1-\alpha$, so the intervals are no longer overly conservative. For $k = k_1 = k_0 + 1$, for both $\alpha = 0.05$ and 0.01, there will be a further adjustment, in which $a_{CP}(k, \alpha)$, which is less than $b(0)$, will be replaced by a number just slightly less than $b(0)$ to avoid excess conservatism.

2. Approximations to binomial probabilities and confidence intervals

As of now, the Clopper-Pearson interval endpoints can easily be computed by computers but not necessarily by calculators. In any case one might not want to use them because they remain too conservative even after the adjustments.

We already saw one approximate interval, the plug-in interval, based on a normal approximation. But before continuing we need to recall some facts about approximations of the binomial distribution that are well known and are, or should be, taught in beginning probability courses. If npq is large, then a binomial(n, p) variable X is approximately normal $N(np, npq)$.

2.1. **Poisson probabilities and approximation.** On the other hand if $p \to 0$ and $n \to \infty$ with $np \to \lambda$ and $0 \leq \lambda < \infty$, then the binomial distribution converges to that of a Poisson(λ) variable Y, for which here are notations: for $k = 0, 1, ...$, letting $0^0 := 1$ in case $\lambda = 0$,

$$P(Y = k) = p(k, \lambda) := e^{-\lambda}\lambda^k/k!, \quad P(k, \lambda) := P(Y \leq k) = \sum_{j=0}^{k} p(j, \lambda),$$

$$Q(k, \lambda) \;:=\; P(Y \geq k) \;=\; \sum_{j=k}^{\infty} p(j, \lambda).$$

In R, $P(k, \lambda)$ can be evaluated as ppois(k, λ), where "pois" indicates the specific distribution and "p" indicates the (cumulative) probability distribution function, analogously as for the binomial and other distributions. One could also find $Q(k, \lambda)$ in R as $1-$ ppois$(k-1, \lambda)$.

If $p \to 1$ and $n \to \infty$ with $nq \to \lambda$ then the distribution of $n - X$ converges to that of Y, Poisson(λ). If n is not large, then neither the normal nor the Poisson approximation to the binomial distribution is good. Similarly, as λ becomes large, the Poisson(λ) becomes approximately $N(\lambda, \lambda)$, but if λ is not large, the normal approximation to the Poisson distribution is not good.

2.2. **Three-regime binomial confidence intervals.** In statistics, where p is not known but X is observed, then for valid confidence intervals we need to proceed as follows. For n not large, specifically in this handout for $n \leq 19$, instead of any approximation, one can just list the confidence interval endpoints in a table. I chose for this purpose adjusted Clopper-Pearson intervals, given in Table 2 in the appendix. This choice is not crucial itself, but secure intervals were chosen for the following reason. For small n, individual values of X have substantial probability. So, there will be substantial jumps in coverage probabilities when one crosses an endpoint. If one makes the coverage probabilities equal to $1-\alpha$ on average, then at points just outside of individual intervals, they could be substantially less than $1-\alpha$, which would be undesirable.

Similarly, if λ is not large, then individual values of Y have substantial probability, and it seemed best to use endpoints that give secure coverage probabilities $\geq 1-\alpha$ in the region where they are used. These will be the Poisson analogue of adjusted Clopper-Pearson endpoints and will be given in Table 1 in the appendix.

If n is large enough (here, $n \geq 20$) but the smaller of X and $n-X$ is no larger than $k_1 = k_1(\alpha)$, then it's better to choose endpoints, specifically $a(k)$ if $k \leq k_1$ or $k = n$, and $b(k)$ if $k = 0$ or $k \geq n - k_1$, based on a Poisson approximation, rather than a normal approximation. The binomial endpoints $a(k)$ for $k \leq k_1$ will be the corresponding Poisson endpoints given in Table 1, divided by n.

For other endpoints, we can use a normal approximation method, but which method? There is a competitor to the plug-in intervals for binomial confidence intervals based on a normal approximation.

2.3. **Quadratic confidence intervals.** The *quadratic* or *Wilson* (1927) intervals are defined as follows. Suppose that p in the interval (3) is replaced by the relation

$$(\hat{p}-p)^2 \;\leq\; z_{\alpha/2}^2 p(1-p)/n,$$

so that the variance estimate $\hat{p}\hat{q}$ is replaced by $pq = p(1-p)$ for the variable p. Then one gets an interval estimator $[a_Q, b_Q]$ by letting $a_Q := a_Q(X,n,\alpha) < b_Q := b_Q(X,n,\alpha)$ be the two roots for p of

$$(\hat{p}-p)^2 \;=\; z_{\alpha/2}^2 p(1-p)/n. \tag{6}$$

If $0 < \hat{p} < 1$ then the quadratic $f(p) = (\hat{p}-p)^2 \; - z_{\alpha/2}^2 p(1-p)/n$ satisfies $f(0) > 0$, $f(\hat{p}) < 0$, and $f(1) > 0$, so by the intermediate value theorem, $0 < a_Q < \hat{p} < b_Q < 1$. If $\hat{p} = 0$ then $a_Q = 0 < b_Q < 1$, or if $\hat{p} = 1$ then $0 < a_Q < b_Q = 1$. The roots of (6) are

$$p \;=\; \frac{2X + z^2 \pm z\sqrt{z^2 + 4X\hat{q}}}{2(n+z^2)} \tag{7}$$

with the minus sign for a_Q and the plus sign for b_Q.

One can see that the quadratic interval is approximating binomial probabilities by normal ones for p at the endpoints of the interval, so that the binomial probabilities approximate those in the definition of the Clopper-Pearson interval (4), (5). Whereas, the plug-in interval crudely uses the normal approximation to the binomial at the center $p = \hat{p}$ where the variance $\hat{p}\hat{q}$ may be quite different from pq at one or both endpoints.

2.4. **Conditions for approximation of quadratic by plug-in intervals.** If not only n but $n\hat{p}\hat{q} = X(n-X)/n$ is large enough, the plug-in and quadratic intervals will be approximately the same, so one can use the simpler plug-in interval. Here are some specific bounds.

Let $z := z_{\alpha/2} = 1.96$ for $\alpha = 0.05$ and 2.576 for $\alpha = 0.01$. If the respective endpoints of the two kinds of intervals are within some $\varepsilon > 0$ of each other, then so must be their centers $(a(j)+b(j))/2$, which are $\hat{p} = X/n$ for the plug-in interval and for the quadratic interval, $(2X+z^2)/(2n+2z^2)$ from (7). The distance between the centers is thus bounded by

$$D_1 := \left|\frac{X}{n} - \frac{2X+z^2}{2n+2z^2}\right| = \frac{z^2|2X-n|}{n(2n+2z^2)} < \frac{z^2}{2n}. \tag{8}$$

The distance from the center to either endpoint is $z\sqrt{\hat{p}\hat{q}/n}$ for the plug-in interval and $z\sqrt{z^2+4X\hat{q}}/(2n+2z^2)$ for the quadratic interval by (7). The absolute difference between these is

$$D_2 = z\left|\frac{(n+z^2)\sqrt{4n\hat{p}\hat{q}} - n\sqrt{4n\hat{p}\hat{q}+z^2}}{2n(n+z^2)}\right|.$$

For any $A > 0$ and $B > 0$, $\sqrt{A} < \sqrt{A+B} < \sqrt{A} + B/(2\sqrt{A})$ by the mean value theorem and since the derivative $(d/dx)\sqrt{x}$ is decreasing. (The bound is most useful for $B << A$.) It follows that we can write $\sqrt{4n\hat{p}\hat{q}+z^2}$ as $\sqrt{4n\hat{p}\hat{q}} + \theta z^2/(4\sqrt{n\hat{p}\hat{q}})$ where $0 < \theta < 1$, and then that

$$D_2 \leq z^3 n^{-3/2} \max(\sqrt{\hat{p}\hat{q}}, 1/(8\sqrt{\hat{p}\hat{q}})).$$

The maximum just written is $\leq 1/(4\sqrt{\hat{p}\hat{q}})$, clearly for the second term, and for the first term, because $p(1-p) \leq 1/4$ for $0 \leq p \leq 1$, attained at $p = 1/2$ only. It follows that $D_2 \leq z^3/(4n\sqrt{n\hat{p}\hat{q}})$. From this and (8), the distance between corresponding endpoints of the quadratic and plug-in intervals is bounded above by

$$D_1 + D_2 \leq \frac{z^2}{2n}\left(1 + \frac{z}{2\sqrt{n\hat{p}\hat{q}}}\right). \tag{9}$$

For $\alpha = 0.05$, taking $z = 1.96$, it will be assumed that

$$n\hat{p}\hat{q} = X(n-X)/n \geq 9,$$

which is equivalent to $\sqrt{n\hat{p}\hat{q}} \geq 3$ and implies that $X \geq 9$ and $n - X \geq 9$, and so that a normal approximation is applicable (for $X \leq 8$ or $X \geq n-8$ I'm recommending Poisson approximations). It follows then that given n, the differences between endpoints are bounded above by $D_1 + D_2 \leq f(z)/n$ where $f(z) = (z^2/2)(1+(z/6)) \leq 2.5483$ using (9) and $\sqrt{n\hat{p}\hat{q}} \geq 3$. We thus have $D_1 + D_2 \leq 10^{-m}$ for $X(n-X)/n \geq 9$ and $n \geq 2.55 \cdot 10^m$, to be applied for $m = 2, 3$, and 4. One wants at least two decimal places of accuracy in the endpoints in nearly any application (for example, political polls, which have other errors of that order of magnitude or more), and no more than 4 places seem to make sense here, where 4 places are given in the tables.

Similarly for $\alpha = 0.01$, taking $z = 2.576$, we'll assume that

$$n\hat{p}\hat{q} = X(n-X)/n \geq 15,$$

which is equivalent to $\sqrt{n\hat{p}\hat{q}} \geq \sqrt{15}$ and implies that $X \geq 15$ and $n - X \geq 15$. For $\min(X, n-X) \leq 14$, a Poisson approximation would be used. Given n, the differences between endpoints are bounded above by $D_1 + D_2 \leq g(z)/n$ where $g(z) = (z^2/2)(1 + \{z/(2\sqrt{15})\}) \leq 4.4213$ using (9) and $\sqrt{n\hat{p}\hat{q}} \geq \sqrt{15}$. We thus have $D_1 + D_2 \leq 10^{-m}$ for $X(n-X)/n \geq 15$ and $n \geq 4.43 \cdot 10^m$, to be applied for $m = 2, 3$, and 4.

So, for example, sufficient conditions for the endpoints of the plug-in and quadratic 99% confidence intervals to differ by at most 0.0001 are that $X(n-X)/n \geq 15$ and

$n \geq 44,300$. If these conditions hold, there is no need to find the quadratic interval, one can just use the plug-in interval.

2.5. **Brown et al.'s comparisons; an example.** The papers by Brown, Cai and DasGupta (2001, 2002) show that the coverage probabilities for various approximate 95% confidence intervals vary and may be quite different from 0.95, not only when p approaches 0 or 1. They show that the quadratic interval, which they (2001) call the *Wilson* interval since apparently Wilson (1927) first discovered it, is distinctly superior to the plug-in interval in its coverage properties.

The problems with the plug-in interval are by no means limited to p close to 0 or 1:

Example. As Brown, Cai and DasGupta (2001, p. 104, Example 2) point out, for $p = 0.5$, presumably the nicest possible value of p, for which the distribution is symmetric, and $n = 40$, the coverage probability of the 95% plug-in interval is 0.919, in other words the probability of getting an interval not containing 0.5 is larger than 0.08 as opposed to the desired 0.05. Let's look at this case in more detail. When $X = 14$, the right endpoint of the plug-in 95% confidence interval is

$$0.35 + 1.96\sqrt{0.35(0.65)/40} = 0.49781 < 0.5.$$

By symmetry since $p = 0.5$, if $X = 26$, the left endpoint of the plug-in 95% confidence interval is $1 - 0.49781 = 0.50219 > 0.5$, so 0.5 is included in the plug-in interval only for $15 \leq X \leq 25$. The probability that $X \leq 14$ is $B(14, 40, 0.5) = 0.040345$ and symmetrically the probability that $X \geq 26$ is $E(26, 40, 0.5) = 0.040345$, so the coverage probability $\kappa(1/2)$ of the plug-in interval in this case is $1 - 2(0.040345) \doteq 0.9193$, confirming Brown et al.'s statement. For the Clopper-Pearson confidence intervals, still for $n = 40$, if $X = 14$ the right endpoint of the interval is 0.51684. For the quadratic interval, it's 0.5049. So these intervals both do contain 0.5, while if $X = 13$ they don't. We have $B(13, 40, 0.5) = E(27, 40, 0.5) = 0.01924$. So the coverage probability of the Clopper-Pearson and quadratic intervals when $n = 40$ and $p = 0.5$ are both $1 - 2(0.01924) \doteq 0.9615$. This coverage probability is closer to the target value of 0.95 by a factor of about 3 relative to the plug-in interval. Also, it may be preferable to have coverage probability a little larger than the target value than to have it smaller.

This is just one case, but it illustrates how the quadratic interval is estimating variance from a value of p at its endpoint, namely 0.5049, which is close to 0.5, the true value. And this is not only by coincidence, but because 14 is the smallest value of X for which the Clopper-Pearson interval contains 0.5, so we'd like the confidence interval to contain 0.5 but not by a wide margin. Whereas, to estimate variance via plug-in, using $p = 0.35$, gives too small a value, and the interval around 0.35 isn't wide enough to contain 0.5. Then the coverage probability is too small.

Brown et al. (2002, Fig. 2) show that for nominal $\alpha = 0.01$ and $n = 30$, the coverage probability of the plug-in interval is strictly less than $1-\alpha = 0.99$ for all p and oscillates wildly to much lower values as $\min(p, 1-p)$ becomes small, e.g. < 0.15.

Another strange and undesirable property of the plug-in interval is that for any $\alpha < 0.3$ and all n large enough, $a(1) < 0 = a(0)$. Specifically, for the 95% plug-in interval with $n \geq 2$ we will have $a(1) < 0$ and $b(n-1) > 1$.

3. Desiderata for interval estimators of p

Some properties generally considered desirable for interval estimators $[a(X), b(X)]$ of the binomial p (e.g. Blyth and Still, 1983; Brown et al. 2001, 2002), are as follows:

1. *Equivariance.* For any $X = 0, 1, ..., n$, $a(n - X) = 1 - b(X)$, $b(n - X) = 1 - a(X)$. All intervals mentioned in this handout are equivariant.

2. *Monotonicity.* $a(X)$ and $b(X)$ should be nondecreasing (preferably strictly increasing) functions of X and nonincreasing (preferably strictly decreasing for $X > 0$) functions of n.

We saw that the 95% (or higher) plug-in interval is not monotone when $n \geq 2$. The other intervals mentioned are all monotone.

3. *Union.* We have $\bigcup_{j=0}^{n}[a(j), b(j)] = [0, 1]$.

If the union doesn't include all of $[0, 1]$ there is some p for which $\kappa(p) = 0$, which seems clearly bad, but this doesn't seem to occur for any commonly used intervals. On the other hand for $n \geq 2$ the 95% plug-in intervals extend below 0 and beyond 1 and so violate the union assumption in a different way.

The remaining desiderata are less precise and can conflict with one another. It's desirable that the coverage probabilities should be close to the nominal $1 - \alpha$. Let's separate this into two parts:

4. *Minimum Coverage.* The minimum coverage probability should not be too much less than $1 - \alpha$.

The intervals to be given in the algorithm in the Appendix have $\kappa(p) \geq 1 - 1.6\alpha$ for $\alpha = 0.05$ or 0.01 and all n and p (not yet proved, but found in computer searches).

5. *Average coverage.* The average coverage probability, namely $\int_0^1 \kappa(p, a(\cdot), b(\cdot))dp$, should be close to $1 - \alpha$ for n large enough.

6. *Shortness.* Consistently with good coverage, the intervals should be as short as possible.

7. *Ease of use and computation.* Intervals proposed to be taught and given in textbooks should not be too complicated or difficult to compute.

The quadratic interval is a little harder to calculate than the plug-in interval, but really not hard to compute, so its improved accuracy is well worth the calculation. An even more easily computed interval is that of Agresti and Coull (1998). It's the modification of the plug-in interval in which n is replaced by $\tilde{n} = n + 4$ and $\hat{p}$ by $(X + 2)/\tilde{n}$, i.e. as if two more successes and two more failures are added to those actually observed. From the comparisons by Brown et al. (2001), the Agresti-Coull interval appears to have minimum coverage probabilities not much less than the nominal ones and tends to be secure for small $\min(p, q)$. Its average coverage probability exceeds the nominal one, with a slowly decreasing difference (bias). The Agresti-Coull intervals tend to be longer than those of some competing intervals whose average coverage probabilities are closer to the nominal.

REFERENCES

Agresti, A., and Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *Amer. Statistician* **52**, 119-126.

Anderson, T. W., and Samuels, S. M. (1967). Some inequalities among binomial and Poisson probabilities. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* (1965), **1**, 1-12. Berkeley and Los Angeles, Univ. of California Press.

Bickel, P. J., and Doksum, K. A. (2001). *Mathematical Statistics: Basic Ideas and Selected Topics*, Vol. I, Second Ed. Prentice-Hall, Upper Saddle River, NJ.

Blyth, C. R., and Still, H. A. (1983). Binomial confidence intervals. *J. Amer. Statist. Assoc.* **78**, 108-116.

Brown, L. D., Cai, T. T., and DasGupta, A. (2001). *Interval estimation for a binomial proportion. Statistical Science* **16**, 101-133. With Comments and a Rejoinder.

Brown, L. D., Cai, T. T., and DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Ann. Statist.* **30**, 160-201.

Cheng, Tseng Tung (1949). The normal approximation to the Poisson distribution and a proof of a conjecture of Ramanujan. *Bull. Amer. Math. Soc.* **55**, 396-401.

Clopper, C. J., and Pearson, E. S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404-413.

Dudley, R. M. (2009). Exposition of T. T. Cheng's Edgeworth approximation of the Poisson distribution, improved. Unpublished manuscript.

Feller, W. (1945). On the normal approximation to the binomial distribution. *Ann. Math. Statist.* **16**, 319-329.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, vol. 1, 3d ed. Wiley, New York.

Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. *J. Amer. Statist. Assoc.* **27**, 713-721.

Jogdeo, K., and Samuels, S. M. (1968). Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation. *Ann. Math. Statist.* **39**, 1191-1195.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *J. Amer. Statist. Assoc.* **22**, 209-212.

4. Appendix: Algorithm and Tables

The proposed algorithm for finding $100(1-\alpha)\%$ confidence intervals $[a(X), b(X)]$ for the binomial p when X successes are observed in n trials and $\alpha = 0.05$ or 0.01 is as follows.

1. If $n \geq 20$, go to step 2. If $n \leq 19$, use the (adjusted, cf. Subsec. 1.4, Clopper-Pearson) intervals given in Table 2.

2. If $n \geq 20$, then use the hybrid endpoints $a_H(X), b_H(X)$ defined as follows: if $\min(X, N-X) \leq k_1(\alpha) = 8$ for $\alpha = 0.05$ and 14 for $\alpha = 0.01$, then go to step 3. If $\min(X, n-X) > k_1(\alpha)$ then use the quadratic endpoints $a_Q(X), b_Q(X)$, specifically, letting $z := z_{\alpha/2}$, and recalling $\hat{p} = X/n$ and $\hat{q} = 1 - \hat{p}$, given by

$$p = \frac{2X + z^2 \pm z\sqrt{z^2 + 4X\hat{q}}}{2(n+z^2)}, \tag{10}$$

where $\pm$ is $-$ for $a_Q(X)$ and $+$ for $b_Q(X)$. For $\alpha = 0.05$, $z \doteq 1.96 \doteq 1.959964$ and $z^2 \doteq 3.84 \doteq 3.84146$, and for $\alpha = 0.01$, $z \doteq 2.576 \doteq 2.5758293$, $z^2 \doteq 6.635 \doteq 6.6348966$.

3. If $\min(X, n-X) \leq k_1(\alpha)$, recalling $k_1(0.05) = 8$ and $k_1(0.01) = 14$, and still for $n \geq 20$: if $0 \leq X \leq k_1(\alpha)$ let $a_H(X) = a_{H,P}(X)/n$ where the hybrid Poisson endpoints $a_{H,P}$ are given in Table 1. Likewise if $X \geq n - k_1(\alpha)$ let $b_H(X) = 1 - a_{H,P}(n-X)/n$. In particular $b_H(n) = 1$.

Let $b_H(0) = z^2_{\alpha/2}/n$, recalling that $z_{0.025} \doteq 1.96$ and $z_{0.005} \doteq 2.576$. Symmetrically let $a_H(n) = 1 - z^2_{\alpha/2}/n$.

Define $a_H(X) = a_Q(X)$ as given by (10) in all other cases, namely if $k_1(\alpha) < X < n$, and $b_H(X) = b_Q(X)$ for $0 < X < n - k_1(\alpha)$.

Table 1. Poisson hybrid left endpoints

k	$a_{H,P}(k, 0.01)$	$a_{H,P}(k, 0.05)$
0	0.0000	0.0000
1	0.0101	0.0513
2	0.1486	0.3554
3	0.4360	0.8177
4	0.8232	1.3663
5	1.2791	1.9701
6	1.7853	2.6130
7	2.3302	3.2853
8	2.9061	3.8415
9	3.5075	
10	4.1302	
11	4.7712	
12	5.4282	
13	6.0991	
14	6.6357	

Table 2, $n \leq 19$. Use $a(k) \equiv 1 - b(n-k)$, $b(k) \equiv 1 - a(n-k)$ for $k > n/2$.

α		0.05		0.01	
n	k	$a(k)$	$b(k)$	$a(k)$	$b(k)$
1	0	.0000	.9500	.0000	.9900
2	0	.0000	.8419	.000	.9293
	1	.0253	.9747	.005	.9950
3	0	.000	.7076	.0000	.8290
	1	.017	.8646	.0033	.9411
4	0	.0000	.6024	.0000	.7341
	1	.0127	.7514	.0025	.8591
	2	.0976	.9024	.0420	.9580
5	0	.0000	.5218	.0000	.6534
	1	.0102	.6574	.0020	.7779
	2	.0764	.8107	.0327	.8944
6	0	.0000	.4593	.0000	.5865
	1	.0085	.6412	.0017	.7057
	2	.0628	.7772	.0268	.8269
	3	.1532	.8468	.0847	.9153
7	0	.0000	.4096	.0000	.5309
	1	.0073	.5787	.0014	.6434
	2	.0534	.7096	.0227	.7637
	3	.1288	.8159	.0708	.8577
8	0	.0000	.3694	.0000	.4843
	1	.0064	.5265	.0013	.6315
	2	.0464	.6509	.0197	.7422
	3	.1111	.7551	.0608	.8303
	4	.1929	.8071	.1210	.8790
9	0	.0000	.3363	.0000	.4450
	1	.0057	.4825	.0011	.5850
	2	.0410	.6001	.0174	.6926
	3	.0977	.7007	.0533	.7809
	4	.1688	.7880	.1053	.8539
10	0	.0000	.3085	.0000	.4113
	1	.0051	.4450	.0010	.5443
	2	.0368	.5561	.0155	.6482
	3	.0873	.6525	.0475	.7351
	4	.1500	.7376	.0932	.8091
	5	.2224	.7776	.1504	.8496
11	0	.0000	.2849	.0000	.3822
	1	.0047	.4128	.0009	.5086
	2	.0333	.5178	.0141	.6085
	3	.0788	.6097	.0428	.6933
	4	.1351	.6921	.0837	.7668
	5	.1996	.7662	.1344	.8307
12	0	.0000	.2646	.0000	.3569
	1	.0043	.3848	.0008	.4770
	2	.0305	.4841	.0128	.5729
	3	.0719	.5719	.0390	.6552
	4	.1229	.6511	.0759	.7275
	5	.1810	.7233	.1215	.7915
	6	.2453	.7547	.1746	.8254
13	0	.0000	.2471	.0000	.3347
	1	.0039	.3603	.0008	.4490
	2	.0281	.4545	.0118	.5410
	3	.0660	.5381	.0358	.6206
	4	.1127	.6143	.0695	.6913
	5	.1657	.6842	.1108	.7546
	6	.2240	.7487	.1588	.8113
14	0	.0000	.2316	.0000	.3151
	1	.0037	.3387	.0007	.4240
	2	.0260	.4281	.0110	.5123
	3	.0611	.5080	.0331	.5892
	4	.1040	.5810	.0640	.6579
	5	.1527	.6486	.1019	.7201
	6	.2061	.7114	.1457	.7766
	7	.2304	.7696	.1947	.8053
15	0	.0000	.2180	.0000	.2976
	1	.0034	.3195	.0007	.4016
	2	.0242	.4046	.0102	.4863
	3	.0568	.4809	.0307	.5605
	4	.0967	.5510	.0594	.6273
	5	.1417	.6162	.0944	.6882
	6	.1909	.6771	.1346	.7439
	7	.2127	.7341	.1795	.7949
16	0	.0000	.2059	.0000	.2819
	1	.0032	.3023	.0006	.3814
	2	.0227	.3835	.0095	.4628
	3	.0531	.4565	.0287	.5344
	4	.0903	.5238	.0554	.5991
	5	.1321	.5866	.0878	.6585
	6	.1778	.6457	.1251	.7132
	7	.1975	.7012	.1665	.7638
	8	.2465	.7535	.2117	.7883
17	0	.0000	.1951	.0000	.2678
	1	.0030	.2869	.0006	.3630
	2	.0213	.3644	.0090	.4413
	3	.0499	.4343	.0269	.5104
	4	.0846	.4990	.0519	.5732
	5	.1238	.5596	.0822	.6310
	6	.1664	.6167	.1168	.6846
	7	.1844	.6708	.1552	.7344
	8	.2298	.7219	.1971	.7807
18	0	.0000	.1853	.0000	.2550
	1	.0028	.2729	.0006	.3463
	2	.0201	.3471	.0085	.4217
	3	.0470	.4142	.0254	.4884
	4	.0797	.4764	.0488	.5492
	5	.1164	.5348	.0772	.6055
	6	.1563	.5901	.1096	.6579
	7	.1730	.6425	.1454	.7068
	8	.2153	.6924	.1844	.7526
	9	.2602	.7398	.2263	.7737
19	0	.0000	.1765	.0000	.2434
	1	.0027	.2603	.0005	.3311
	2	.0190	.3314	.0080	.4037
	3	.0445	.3958	.0240	.4682
	4	.0753	.4557	.0461	.5271
	5	.1099	.5120	.0728	.5818
	6	.1475	.5655	.1032	.6329
	7	.1629	.6164	.1368	.6809
	8	.2025	.6650	.1733	.7260
	9	.2445	.7114	.2124	.7684