18.443, February, 2008

CONFIDENCE INTERVALS FOR PROPORTIONS

We'll define four different kinds of confidence intervals for proportions: "exact binomial," and three normal approximations, "plug-in," "margin-of-error" (simplest) and "quadratic" (requires a little more calculation, tends to be the most accurate of the three approximations).

For differences of two proportions, only the plug-in and margin-of-error intervals are defined.

A main point to keep in mind is that binomial probabilities are well approximated by normal probabilities if n is large and p is not too close to 0 or 1. If p is close to 0 then the normal approximation fails and a Poisson approximation is better. If p is close to 1, then we can consider the number of failures rather than successes and it will have approximately a Poisson distribution.

If an endpoint of a confidence interval resulting from a normal approximation approaches 0 or 1, then you should be aware that the approximation may not be valid and just say so. Further calculations are not required. (In the real world, you should use a computer to get exact binomial intervals.) If the endpoints stay away from 0 or 1 and n is reasonably large, then you should say what kind of interval you are using.

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

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

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

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

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

Then some identities for these probabilities are

$$B(k, n, p) \ \equiv \ 1 - E(k + 1, n, p), \qquad B(k, n, p) \ \equiv \ E(n - k, n, q).$$

Recall that when we have a $100(1-\alpha)\%$ confidence interval $[\mu_\ell, \mu_r]$ for the unknown mean μ of a normal distribution, based on an observed sample mean $\overline{X}$ and either normal quantiles if σ^2 is known or t quantiles if σ^2 has to be estimated from the data, it means that if the true value of μ were μ_ℓ, then the probability of observing an $\overline{X}$ as large or larger than the actual value is $\alpha/2$, and likewise, if the true value of μ were μ_r, then the probability of observing an $\overline{X}$ as small or smaller than the actual value is $\alpha/2$.

For the binomial case, $\overline{X} = Y/n$ is the maximum likelihood estimate (Rice, p. 314, stated in a problem 6a) and an unbiased estimator of the unknown p, and we want to find confidence intervals for p. Different confidence intervals will be defined and compared.

If $0 < Y < n$, the *exact binomial* $100(1-\alpha)\%$ confidence interval for p will be an interval $[p_\ell, p_r]$ such that if the true p were p_ℓ, there would be a probability $\alpha/2$ of observing an $\overline{X}$ as large or larger than the actual one, in other words,

$$(1) \qquad E(Y, n, p_\ell) \ = \ \alpha/2.$$

Similarly, if the true p were p_r, there would be a probability $\alpha/2$ of observing an $\overline{X}$ as small or smaller than the actual one, in other words,

$$(2) \qquad B(Y, n, p_r) \ = \ \alpha/2.$$

We know that if n is large and p is not too close to 0 or 1, then the binomial (n, p) distribution can be well approximated by a normal distribution with the same mean and variance. Thus $\overline{X}$ will be approximately normal with mean p, variance pq/n and so standard deviation $\sqrt{pq/n}$. But we have the problem that p and $q = 1 - p$ are unknown. If we have an estimate for p, say $\hat{p}$, such as the MLE $\hat{p} = \overline{X}$, and $\hat{q} = 1 - \hat{p}$, we can consider the approximate confidence interval

$$(3) \qquad [\hat{p} - z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}, \hat{p} + z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}].$$

Recall that z_β is defined so that if Z has a $N(0,1)$ distribution then $P(Z \geq z_\beta) = \beta$. When $\hat{p} = \overline{X}$, we will refer to the interval defined by (3) as a *plug-in* confidence interval and call it $[\overline{p}_\ell, \overline{p}_r]$. Now, notice that in (1) and (2), the probabilities p_ℓ, p_r and therefore the variances $p_\ell(1-p_\ell)$, $p_r(1-p_r)$ are different in general. Thus, the plug-in confidence interval works well if n is large enough and $\overline{X}$ is far enough from 0 or 1 so that not only the normal approximations to the binomial probabilities work well but also we have approximate equality of variances,

$$(4) \qquad \overline{p}_\ell(1-\overline{p}_\ell) \sim \overline{X}(1-\overline{X}) \sim \overline{p}_r(1-\overline{p}_r).$$

We can also notice that $pq \leq 1/4$ always and pq is close to $1/4$ for p in an interval around $1/2$, such as 0.4 to 0.6. So a quick and simple approach, especially suitable for p in that range, is to use (3) with the variance estimate $\hat{p}\hat{q}$ replaced by $1/4$, as in the example of political polling, in an election between two fairly evenly matched candidates. Since political poll results are presented in terms of a "margin of error" we will call the approximate confidence intervals

$$[\overline{X} - z_{\alpha/2}/(2\sqrt{n}), \overline{X} + z_{\alpha/2}/(2\sqrt{n})]$$

margin-of-error confidence intervals.

Suppose that we write the confidence interval (3) in the form

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

so that the variance estimate $\hat{p}\hat{q}$ is replaced by $pq = p(1-p)$ for the variable p. Then we can get an approximate confidence interval $[\tilde{p}_\ell, \tilde{p}_r]$ by letting $\tilde{p}_\ell < \tilde{p}_r$ be the two roots of the quadratic equation

$$(\overline{X} - p)^2 \;=\; z^2_{\alpha/2} p(1-p)/n.$$

We call $[\tilde{p}_\ell, \tilde{p}_r]$ a *quadratic* confidence interval. These intervals allow for different variances for the different endpoints of the confidence interval, so they do not require a condition like (4). Still, they require that the normal approximation of binomial probabilities should be valid for the p's at both ends of the interval. For the quadratic intervals, n only needs to be moderately large, compared to the larger n's needed for (4). Also, the confidence interval should not approach too close to the endpoints $p = 0$ or 1.

For differences of two proportions, suppose we have Y_1 with a binomial (n_1, p_1) distribution and Y_2 with a binomial (n_2, p_2) distribution. A corresponding margin-of-error confidence interval can be gotten by replacing the estimates of the variances $p_1(1-p_1)$ and $p_2(1-p_2)$ both by $1/4$. Now instead of $1/\sqrt{n}$ we will have $\sqrt{(n_1+n_2)/(n_1 n_2)}$. "Exact binomial" and "quadratic" confidence intervals are not available in this case. The plug-in intervals only work well when n_1 and n_2 are fairly large and (4) holds for both binomial variables. The margin-of-error intervals are all right if the estimates of p_1 and p_2 are both in the interval $[0.4, 0.6]$, otherwise these intervals may be wider than necessary.

If n is not very large or the estimate of p_1 or p_2 is close to 0 or 1, then we can't recommend any good confidence intervals for $p_1 - p_2$.

Now let's consider some examples of confidence intervals for one p. First, suppose $\overline{X} = 0$ or 1, in other words $Y = 0$ or n. Then the plug-in estimate of the variance is 0 and the plug-in confidence interval has 0 width. We can let α approach 0 so $z_{\alpha/2}$ gets very large, but the same still happens. Does this mean if we observe $Y = 0$, we are sure $p = 0$? No, this is an illustration of how badly the plug-in interval can behave when $\overline{X}$ approaches 0 or 1.

In case $Y = 0$, the exact binomial confidence interval for p is one-sided and easy to compute, being $[0, p_{r,0}]$ where $(1 - p_{r,0})^n = \alpha$ (not $\alpha/2$). Likewise if $Y = n$, the exact binomial confidence interval is $[p_{\ell,1}, 1]$ where $p^n_{\ell,1} = \alpha$.

In the next two examples, Poisson approximations will be mentioned for comparison. They still require use of a computer and are only approximations, so we didn't emphasize them.

A sample problem: if 5 of 100 rear view mirrors are defective, find a 95% confidence interval for the probability p that a mirror is defective. The confidence intervals are:

Exact binomial: $[0.01643, 0.11283]$

Poisson approx.: $[0.0162, 0.117]$

Plug-in interval: $[0.007, 0.093]$

Margin-of-error: $[-0.048, 0.148]$

Quadratic: $[0.0215, 0.112]$.

The margin-of-error interval is much too wide at both ends. It uses a variance that's too big in this case.

The plug-in interval is the next-worst. It's too wide at the lower end, too narrow at the upper end. In the original problem it was said that formerly 10% of mirrors were defective and an experiment was done to see if the mirrors were now better. The plug-in interval would give "confidence" that they were, but the more exact intervals don't.

The quadratic interval has a good upper endpoint. This illustrates that it gets the variance correct, using different variances for the two endpoints, and the upper endpoint is in a range where the normal approximation works better. We see that for the lower endpoint the Poisson approximation is much better than any of the three normal approximations.

For another example, let $n = 642$ and $Y = 24$. Then

Exact binomial interval: $[0.0241, 0.0551]$,

Poisson approx.: $[0.0240, 0.0556]$

Plug-in interval: $[0.0227, 0.0521]$,

Quadratic interval: $[0.0252, 0.0550]$.

Here the Poisson approximation works well. The normal approximate intervals are not bad either. In this case, although $\overline{X}$ is fairly close to 0, apparently n is large enough to make normal approximations work fairly well. The quadratic interval does quite well again at its upper endpoint where the normal approximation works better.

An article showing the general superiority of the quadratic interval (which they call the "Wilson" interval) is "Interval estimation for a binomial proportion" by L. D. Brown, T. T. Cai, and A. DasGupta, *Statistical Science* **16** (2001), pp. 101-133.