COVERAGE PROBABILITIES AND CONFIDENCE INTERVALS

This handout relates to the material of problem sets 1 and 2 rather than ps3 (it doesn't relate to regression, although often people do define confidence intervals relating to regression).

Let θ be an unknown parameter, which may be a vector, such as (μ, σ) for a normal distribution, or a scalar such as the probability p in a binomial distribution. Suppose given a vector $X = (X_1, ..., X_n)$ of observations where X_j are i.i.d. with a distribution of a given form having parameter θ.

Recall that a *statistic* is a function $T = T(X)$ of the observations, and suppose there's a real-valued function $g(\theta)$ that we'd like to estimate. An *estimator* of $g(\theta)$ must be a statistic, but to have a precise definition we need to give some property of the estimator such as being unbiased or a maximum likelihood estimator. So for now let's just say that an estimator (of $g(\theta)$) must be a statistic.

Similarly, an *interval estimator* will be defined as a pair of real-valued statistics $a(X)$ and $b(X)$ such that $a(X) \leq b(X)$ for all X. Here we have in mind using the interval $[a(X), b(X)]$ as a confidence interval for $g(\theta)$, but that turns out also to be hard to define precisely in general. The notation $a(\cdot)$ will be used for the statistic (function) whose value at any X is $a(X)$, whereas $a(X)$ would be a particular value of the statistic.

Let P_θ denote probability when θ is the true value of the parameter. The *coverage probability* for a given interval estimator $[a(\cdot), b(\cdot)]$, function $g(\cdot)$ of θ, and a given θ is defined as

$$\kappa(\theta, g(\theta), a(\cdot), b(\cdot)) = P_\theta[a(X) \leq \theta \leq b(X)]. \tag{1}$$

Many texts define $1 - \alpha$ or $100(1 - \alpha)\%$ confidence intervals for $g(\cdot)$ as interval estimators $[a(\cdot), b(\cdot)]$ whose coverage probability for $g(\theta)$ equals $1 - \alpha$ for all θ, perhaps only approximately.

For the mean μ of a normal distribution, there are well-known interval estimators based on the t distribution such that whatever μ and σ^2 are, the coverage probability exactly equals $1-\alpha$. These are usually taken to be two-sided, with $-\infty < a(X) < b(X) < +\infty$ for all X, as we did. It's also possible to consider 1-sided intervals $(-\infty, b(\cdot)]$ or $[a(\cdot), +\infty)$. Likewise, there are interval estimators for the variance σ^2 or standard deviation σ of a normal distribution, based on the χ^2 distribution, having exact coverage probability $1 - \alpha$ for all μ and σ, as we've seen. So for the parameters of normal distributions there is general agreement on what is a "$100(1 - \alpha)\%$ confidence interval."

However, for the binomial p, the situation is different. Not only are the usual confidence intervals only approximations, but even in case we use "exact" intervals, the coverage probability in general does not equal $1-\alpha$. The paper by Brown, Cai and DasGupta (2001) shows that when one aims at $\alpha = 0.05$ (95% confidence) for $n = 50$, the coverage probabilities for various approximate confidence intervals vary and may be quite different from 0.95, especially when p approaches 0 or 1.

As an example let's consider the case where p is close to 1, specifically $p^n > 1 - \alpha$ with $\alpha < 1/2$. Here p^n is the probability that the binomial random variable $S_n = n$, and when that happens, we've defined the "exact" $100(1-\alpha)\%$ confidence interval to be the 1-sided interval $[\alpha^{1/n}, 1]$. Our p will be in this interval since $p^n > 1-\alpha > \alpha$. Thus the coverage probability is *at least* p^n, which is larger than $1-\alpha$.

For given values of n and α, the confidence interval ("exact" or approximate) when $S_n = n$ must be some interval depending only on α and n. Suppose it's an interval $[a(n, \alpha), 1]$ with $a(n, \alpha) < 1$. Then, if p is larger than $a(n, \alpha)$ and also larger than $(1-\alpha)^{1/n}$, the coverage probability for such a p will be at least p^n which is larger than $1-\alpha$. So, there is *no way* to define confidence intervals for binomial probabilities such that the coverage probabilities always equal $1-\alpha$.

We can, however, prefer approximate confidence intervals whose coverage probabilities are reasonably close to $1-\alpha$ for p in as wide a range as possible (but not too close to 0 or 1). The numerical investigation by Brown, Cai and DasGupta (2001) shows that the quadratic interval is superior to the plug-in interval in this respect.

The problems with the plug-in interval are by no means limited to p close to 0 or 1. As Brown, Cai and DasGupta (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 $S_n = 14$, 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 $S_n = 26$, 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 $15 \leq S_n \leq 25$. The probability that $S_n \leq 14$ is $B(14, 40, 0.5) = 0.040345$ and symmetrically the probability that $S_n \geq 26$ is $E(26, 40, 0.5) = 0.040345$, so the coverage probability of the plug-in interval in this case is $1 - 2(0.040345) \doteq 0.9193$, confirming Brown et al.'s statement.

For the "exact" confidence intervals defined in the handout "Confidence intervals for proportions," still for $n = 40$, if $S_n = 14$ 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 $S_n = 13$ they don't. We have $B(13, 40, 0.5) = E(27, 40, 0.5) = 0.01924$. So the coverage probability of the exact 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 a little 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 S_n for which the exact 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.

Now let's return to the general question of what "confidence interval" means. For $0 < \alpha < 1$, Bickel and Doksum (2001, p. 235) define a $100(1-\alpha)\%$ confidence interval for $g(\cdot)$ as an interval estimator $[a(\cdot), b(\cdot)]$ such that the coverage probability

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

for all possible θ. That is a mathematically satisfying definition in the sense of being precise, but it isn't what most statisticians actually use. Rather, many statisticians would refer to an interval satisfying (2) as a *conservative* $1-\alpha$ confidence interval for $g(\theta)$. They 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. Also, a desirable property of an interval estimator is that the functions $a(\cdot)$ and $b(\cdot)$ be relatively easy to compute.

For the binomial p, Bickel and Doksum (pp. 237-238) recommend the quadratic interval, whose coverage probabilities are usually closer to the target $1-\alpha$ than are those of the plug-in interval, as shown by Brown, Cai and DasGupta. In the special cases where one observes $S_n = n$ or 0 successes, one can use instead the "exact" intervals $[\alpha^{1/n}, 1]$ or $[0, 1-\alpha^{1/n}]$. Bickel and Doksum were evidently alert to the literature, as their book came out in the same year as the paper, evidently they had access to a pre-publication version. The quadratic interval is a little harder to calculate by hand than the classical plug-in interval, but really not hard at all to compute by machine, so its improved accuracy is well worth the calculation.

There's a considerable research literature on confidence intervals for the binomial p. For a given n there are just $n+1$ possible values of $S_n = 0, 1, ..., n$ and so an interval estimator also has just $n+1$ possible pairs of endpoints $a(S_n), b(S_n)$. Among intervals that are actually conservative, one can impose further criteria such as that the sum of the lengths of the $n+1$ intervals be as small as possible. In this course, we won't get into further details.

Bickel and Doksum say in their preface (to both editions) that their book, although an introduction to statistics, is aimed at graduate students who have previously studied probability and linear algebra, but not necessarily measure theory.

REFERENCES

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.

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.