Setting an Upper Confidence Bound on Proportion Nonconforming

By Joel Dobson

Q Is it possible to calculate a 95% upper confidence bound for the proportion nonconforming in the uninspected remaining units in a lot of *N* units, from which we sample *n* of them and find *x* nonconforming?

A Let's start with an example to explain the method and the calculations giving the general equations.

Suppose our lot contains $N = 1{,}000$ units. Among a randomly chosen sample of $n = 100$, we find $x =$ 1 bad unit. For the remaining $N - n = 900$ units, we estimate that 1% will be nonconforming, which is 9, but we can say more than that. In particular, we feel 95% confident that no more than 4.656% of them will be nonconforming, which is about 42 units.

How did we arrive at this percentage and number?

To illustrate the thinking behind the method that follows, let's first try to guess the smallest value of the parameter, *p*, of a binomial distribution with a sample of $n = 100$, which gives us: {the probability that *x* is less than or equal to one} is less than 0.05, or, symbolically, 'Prob($x <= 1$) < 0.05.' Here, the variable *x* is the count of bad units observed among our sample.

We suspect this value of *p* should be larger than the observed 1% nonconforming because we desire Prob($x < = 1$) < 0.05. Alternatively, this is the same as saying we desire Prob($x > 1$) > 0.95, or Prob($x >= 2$) > 0.95. What is the smallest value of *p* that attains this goal? As our first guess, let's try a value of $p = 0.02$. The probability density function, or pdf, of a binomial has this equation:

$$f(x) = \binom{n}{x} * p^x * (1-p)^{(n-x)}$$

where *X* ranges over the whole numbers from 0 to *n*. For our example, we have:

$$f(x) = \binom{100}{x} * 0.02^x * (0.98)^{(100-x)}$$

Or:

$$f(x) = \frac{100!}{x!(100-x)!} * 0.02^x * (0.98)^{(100-x)}$$

The first few *x* values give function values of:

f(0) = 100! / (0! (100-0)!)*0.02^0*0.98^100 = 1 * 0.02^0 * 0.98^100 = 0.132620,

f(1) = 100! /(1! (100-1)!)*0.02^1*0.98^99 = 100 * 0.02^1 * 0.98^99 = 0.270652,

f(2) = 100! /(2! (100-2)!)*0.02^2*0.98^98 = (100*99/2) * 0.02^2 * 0.98^98 = 0.273414,

f(3) = 100!/(3! (100-3)!)*0.02^3*0.98^97 = (100*99*98/6) * 0.02^3 * 0.98^97 = 0.182276,

and so forth. By the time we get to $x = 12$, the value of this pdf function is close to zero.

We can calculate these values in Excel using equations like "=binomdist(0, 100, 0.02, 0)", "=binomdist(1, 100, 0.02, 0)", "=binomdist(2, 100, 0.02, 0)", etc. The probability of seeing 1 or fewer is 0.403272. We can get this from Excel using "=binomdist(1, 100, 0.02, 1)". The probability of seeing 2 or more is 1 - 0.403272 = 0.596728, or about 60%. Figure 1 shows a graph of the pdf (for the first 13 terms). Figure 2 shows a graph of the cumulative distribution function, or CDF (for the first 13 terms).

The CDF at $x = 1$ is 40%, corresponding to the probability that X is less than or equal to one. Because Prob($x > 1$) = 1 - Prob($x <= 1$), we say that 0.02 forms a 60% upper confidence bound on the nonconformance. The upper bound of nonconformance is no more than 0.02, or 2% at 60% confidence. This 60% probability corresponds to the probability that x is greater than one, or that x is greater than or equal to two, if p were 0.02. Using our guess of $p = 0.02$, we have calculated Prob($x < = 1$) < 0.40, or Prob($x > 1$) > 0.60. Our goal was to find the smallest value of p that gives us Prob($x <= 1$) < 0.05, or Prob($x > 1$) > 0.95. Clearly, our first guess of $p = 0.02$ was not large enough. We have only moved 60% probability to the values of x larger than one. Our stated goal was to calculate a 95% upper confidence bound for p.

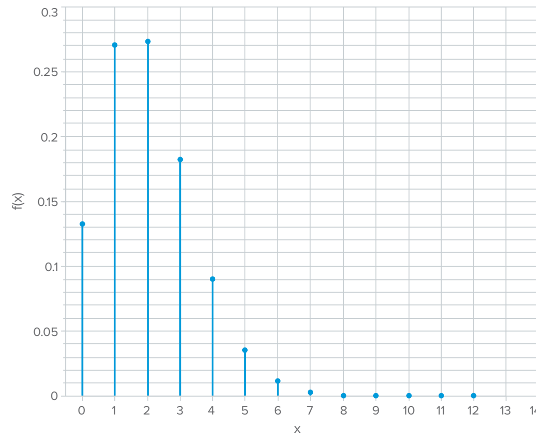
If we use a solver method in Excel, we can estimate the 95% upper confidence bound for the proportion nonconforming to be 0.0465598. We do this by solving binomdist(1, 100, p, 1) = 0.05 for p. This gives the value p = 0.0465598 = 4.65598%. That is to say, binomdist(1, 100, 0.0465598, 1) = 0.05. The smallest value of p which attains our goal of Prob($x > 1$) > 0.95, or Prob($x >= 2$) > 0.95, is p = 0.0465598. Figure 3 shows a revised graph of the pdf (for the first 13 terms). Figure 4 shows a revised graph of the cumulative distribution function, or CDF (for the first 13 terms).

Notice that 95% of the probability occurs for x values larger than or equal to 2, when the value of p is 0.0465598. Prob($x = 2$ or $x = 3$ or ... or $x = 100$) = Prob($x >= 2$) = 0.95. Likewise, five percent of the probability occurs for x values less than or equal to 1. Prob($x = 0$ or $x = 1$) = Prob($x <= 1$) = 0.05. We conclude that a 95% upper confidence bound for the proportion nonconforming is 0.0465598, or 4.65598%.

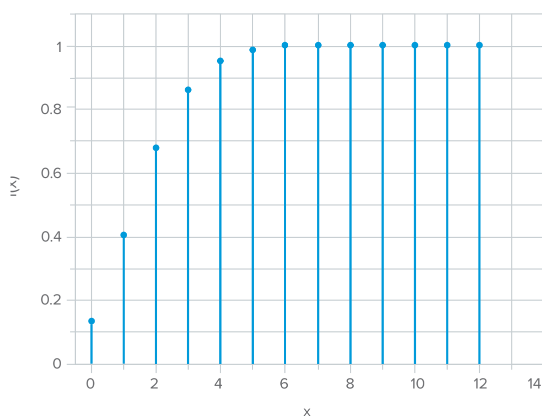
For the remaining 900 units we estimate that 1% will be nonconforming, which is 9 units, but we can say more than that. In particular, we have as our 95% upper confidence bound that no more than 4.656% of them will be nonconforming, which corresponds to approximately 42 units.

In summary, this article illustrates the method to calculate a 95% upper confidence bound for the proportion nonconforming in the remaining ($N - n$) units that were not inspected, when we have found a count of x nonconforming among the random n we sampled. We used a heuristic, verging on didactic, approach to explain how one would allocate the 5% probability and the 95% probability. We estimated the value of the parameter, p, from a binomial distribution using Excel functions and a solver approach. This method forms a valuable tool in the toolbox of the statistician or reliability engineer, and a well-worn tool indeed.

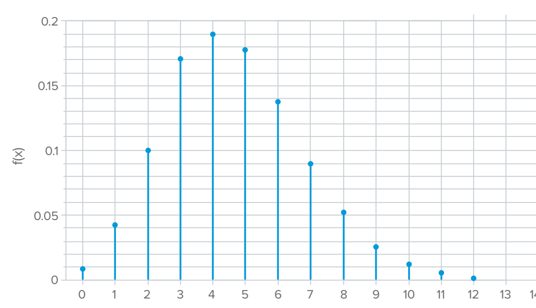
1 A graph of the pdf looks like this (for the first 13 terms).



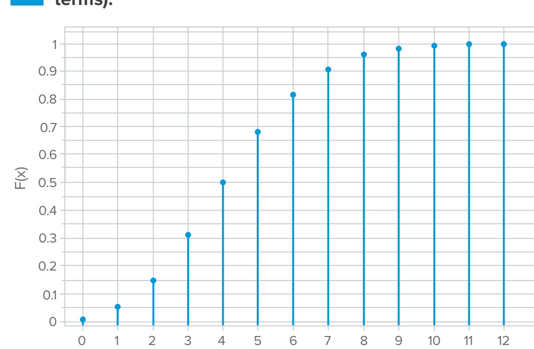
2 A graph of the CDF looks like this (for the first 13 terms).



3 Now the revised pdf graph looks like this (for the first 13 terms).



4 The revised CDF graph looks like this (for the first 13 terms).



Joel Dobson is a senior statistician and data scientist at Texas Instruments, and he is a member of the committee on quality and statistics (E11), where he is a member at large on the executive subcommittee.

Dean V. Neubauer, the Data Points column coordinator, is engineering fellow and chief statistician, Corning Inc. A member at large on the executive subcommittee of the committee on quality and statistics (E11), he is an ASTM International fellow and a past chair of the E11 committee.