Probability and Statistics Notes

Δ

g

Note 10

Some Special Case Results for Confidence and Reliability in a Finite Population

> by Chris Ashley, Ken Chen, and J. Barry

21 May 1981

Air Force Weapons Laboratory

Abstract

In Reference [R1] (among other places) there is derived a result which gives confidence level as a function of reliability of a finite population and experimental data (subject only to the maximum ignorance assumption). The present note derives analytic simplifications of this result for two special cases, viz., uniformly successful testing and uniformly failed testing. It can be extremely easy to calculate exact confidence values when these simplified results apply (see, for example, paragraph 9).

Contents

		Page
Abstract.		1
Preface.		3
Some Special Reliat	Case Results for Confidence and pility in a Finite Population.	4
Appendix A.	Proof of Lemma 1.	14
Appendix B.	Alternative proofs of Lemma 2.	15
Appendix C.	Alternative proofs of Lemma 3.	16
Appendix D.	A direct proof of Corollary 1.	18
References.		19

Preface

In Reference [R1] (among other places) there is derived a result which gives confidence level as a function of reliability of a finite population and experimental data (subject only to the maximum ignorance assumption). Of recurrent special interest is the confidence level when the test results are uniformly good, i.e., all tested elements pass their tests. One situation in which this special case is of interest is optimistic test planning. question in that situation is, "If everything passes the test, what is the smallest sample size we can test which will provide certain predetermined minimum levels of confidence and reliability?" (The reason for the interest in this question is, required sample size is minimized by having all tested elements pass.) Sometimes of special interest also is the confidence level when the test results are uniformly bad, i.e., all tested elements fail. A situation in which this case is of special interest is when one wants to know the smallest sample size which can provide specified minimum confidence in the reliability of a statement that a member of the population won't work. (This situation is essentially the same as that for optimistic test planning, above.) Another situation is that in which the (pessimistic) question is, "If everything tested fails, how many will we have to test before the confidence level or reliability drops below a prespecified threshold?"

To answer such questions it is necessary to be able to invert confidence as a function of reliability, population size, and sample size. To achieve such inversions economically it is desirable that the confidence function be as simple as possible. The present note derives a few such analytic simplifications. The principal results for purposes of applications are equations (11) and (14), and their even-more-special-case corollaries, implications (12) and (15). These results allow some reliability-confidence calculations which previously were practicable only on a computer now to be performed easily on any hand calculator, and sometimes even by hand.

1. (Introduction.) In Reference [R1] there is derived a result which gives confidence level as a function of reliability of a finite population and experimental data (subject only to the maximum ignorance assumption). That result is presented as equation (4) in that reference. In Reference [R2] the same result is presented again, there as equation (1), but with a slight change in notation. With that change it appears as:

$$C(R,N,L,M) = \frac{\sum_{I=\lceil NR \rceil}^{N-L+M} \left[\binom{I}{M} \binom{N-I}{L-M} \right]}{\sum_{I=M}^{N-L+M} \left[\binom{I}{M} \binom{N-I}{L-M} \right]}$$
(1)

where

 $N \stackrel{\triangle}{=}$ finite cardinality of the population,

 $L \stackrel{\triangle}{=}$ number tested to date (sampling done randomly without replacement),

M $\stackrel{\triangle}{=}$ number which have passed the test to date,

R = population reliability (by which we mean a lower bound on the fraction of the population which would pass the test if every member of the population were tested),

 $C \stackrel{\triangle}{=}$ confidence level in R as a result of test data to date,

 $[x] \stackrel{\triangle}{=} least integer greater than or equal to x, and$

 $[x] \stackrel{\triangle}{=} greatest integer less than x.$

2. (Purpose.) The goal of the present note is to present analytic simplifications of equation (1), above, for two special cases, viz., M = L (i.e., all tests successful) and M = 0 (i.e., all tests failed). To achieve this goal we begin by establishing a theorem and three simple lemmas.

3. Symmetry Theorem for Confidence in Reliability of a Finite Population:

$$C(R,N,L,M) = 1 - C(1-R,N,L,L-M)$$
 (2)

Proof:

$$1 - C(1-R,N,L,L-M) \xrightarrow{(1)} 1 - \frac{\sum_{I=\lceil N(1-R)\rceil}^{N-L+(L-M)} \left[\binom{I}{L-M} \binom{N-I}{L-(L-M)} \right]}{\sum_{I=L-M}^{N-M} \left[\binom{I}{L-M} \binom{N-I}{L-(L-M)} \right]} = \frac{N-M}{N-M} \left[\binom{I}{L-M} \binom{N-I}{L-M-M} \right]$$

$$= 1 - \frac{\sum_{I=\lceil N(1-R)\rceil}^{N-M} \left[\binom{I}{L-M} \binom{N-I}{M} \right]}{\sum_{I=L-M}^{N-M} \left[\binom{I}{L-M} \binom{N-I}{M} \right]} =$$

$$= \frac{\sum_{I=L-M}^{N-M} \left[\binom{I}{L-M} \binom{N-I}{M} \right]}{\sum_{I=L-M}^{N-M} \left[\binom{I}{L-M} \binom{N-I}{M} \right]} - \frac{\sum_{I=\lceil N(1-R) \rceil}^{N-M} \left[\binom{I}{L-M} \binom{N-I}{M} \right]}{\sum_{I=L-M}^{N-M} \left[\binom{I}{L-M} \binom{N-I}{M} \right]}$$

$$= \frac{\sum_{I=L-M}^{N-M} \left[\binom{I}{L-M} \binom{N-I}{M} \right] - \sum_{I-\lceil N(1-R) \rceil}^{N-M} \left[\binom{I}{L-M} \binom{N-I}{M} \right]}{\sum_{I=L-M}^{N-M} \left[\binom{I}{L-M} \binom{N-I}{M} \right]}$$

$$= \frac{\sum_{I=L-M}^{\lfloor N(1-R)\rfloor} \left[\binom{I}{L-M} \binom{N-I}{M} \right]}{\sum_{I=L-M}^{N-M} \left[\binom{I}{L-M} \binom{N-I}{M} \right]} \frac{\underbrace{K^{\triangle}=L-M}}{M}$$

$$= 1 - C(1-R,N,L,L-M) \qquad \frac{\sum_{I=K}^{L-M} \left[\begin{pmatrix} I \\ K \end{pmatrix} \begin{pmatrix} N-I \\ L-K \end{pmatrix} \right]}{\sum_{I=K} \left[\begin{pmatrix} I \\ K \end{pmatrix} \begin{pmatrix} N-I \\ L-K \end{pmatrix} \right]} = \frac{\sum_{I=K}^{L-K} \left[\begin{pmatrix} I \\ K \end{pmatrix} \begin{pmatrix} N-I \\ L-K \end{pmatrix} \right]}{\sum_{I=K}^{N-L+K} \left[\begin{pmatrix} I \\ K \end{pmatrix} \begin{pmatrix} N-I \\ L-K \end{pmatrix} \right]}$$

Now, since M is defined in the present note (and in Reference [R2]) as number observed good to date, it follows that K as defined above is the number observed bad to date. Therefore the last expression above is the same as the right hand side of equation (4) in Reference [R1], which is just C(R,N,L,M), q.e.d.

This lemma is just equation 3.1.4 in Reference [R3]. It is easy to remember, since this is the fact which one uses to write down Pascal's Triangle. (The Lth entry in the Nth row of Pascal's Triangle is $\binom{N}{L}$.) A proof is provided as Appendix A of the present note.

5. Lemma 2:
$$0 \le J \longrightarrow \sum_{I=N-l}^{N} {I \choose L} = {N+1 \choose L+1} - {N-J \choose L+1} \qquad (4)$$

Proof: By induction (on J):

Induction step (true for J-1 → true for J):

Assume
$$\sum_{I=N-(J-1)}^{N} {I \choose L} = {N+1 \choose L+1} - {N-(J-1) \choose L+1}$$
 (5)

Then
$$\sum_{I=N-J}^{N} {I \choose L} = {N-J \choose L} + \sum_{I=N-J+1}^{N} {I \choose L} \frac{(5)}{(5)}$$

$$\frac{(5)}{L} {N-J \choose L} + {N+1 \choose L+1} - {N-(J-1) \choose L+1} =$$

$$= {N+1 \choose L+1} - {N-J \choose L+1} - {N-J \choose L} \frac{(3)}{(3)}$$

$$\frac{(3)}{L+1} {N+1 \choose L+1} - {N-J \choose L+1} , q.e.d.$$

Alternative, non-inductive (i.e., direct) proofs of Lemma 2 are also provided, in Appendix B.

6. Lemma 3:
$$\sum_{I=1}^{N} {I \choose L} = {N+1 \choose L+1}$$
 (6) .

Proof: By induction (on N):

Ground step (N=1):

If N=1 and there is a sample then L=1 .

So N=1 means we must prove

$$\sum_{I=1}^{1} \begin{pmatrix} I \\ 1 \end{pmatrix} = \begin{pmatrix} 1+1 \\ 1+1 \end{pmatrix} \dots$$

. . . which is obvious.

Induction step (true for N-1 ⇒ true for N):

Assume
$$\sum_{I=1}^{N-1} {I \choose L} = {N \choose L+1}$$
 (7)

Then
$$\sum_{I-L}^{N} \begin{pmatrix} I \\ L \end{pmatrix} = \begin{pmatrix} N \\ L \end{pmatrix} + \sum_{I=L}^{N-1} \begin{pmatrix} I \\ L \end{pmatrix} \xrightarrow{(7)}$$

$$\frac{(7)}{L} \begin{pmatrix} N \\ L \end{pmatrix} + \begin{pmatrix} N \\ L+1 \end{pmatrix} \xrightarrow{(3)} \begin{pmatrix} N+1 \\ L+1 \end{pmatrix} , q.e.d.$$

Alternative, non-inductive (i.e., direct) proofs of Lemma 3 (from Lemma 2) are also provided, in Appendix C.

7. Returning to paragraph 2, above, we can now derive the special case of equation (1) for M = L, i.e., all tests successful. Well,

$$C(R,N,L,L) = \frac{\sum_{I=\lceil NR \rceil}^{N-L+L} \left[\begin{pmatrix} I \\ L \end{pmatrix} \begin{pmatrix} N-I \\ L-L \end{pmatrix} \right]}{\sum_{I=L}^{N-L+L} \left[\begin{pmatrix} I \\ L \end{pmatrix} \begin{pmatrix} N-I \\ L-L \end{pmatrix} \right]} = \frac{\sum_{I=\lceil NR \rceil}^{N} \begin{pmatrix} I \\ L \end{pmatrix}}{\sum_{I=L}^{N} \begin{pmatrix} I \\ L \end{pmatrix}} = \frac{\sum_{I=\lceil NR \rceil}^{N} \begin{pmatrix} I \\ L \end{pmatrix}}{\sum_{I=L}^{N} \begin{pmatrix} I \\ L \end{pmatrix}} = \frac{\sum_{I=\lceil NR \rceil}^{N} \begin{pmatrix} I \\ L \end{pmatrix}}{\binom{N+1}{L+1}}$$

$$(8)$$

To proceed we may consider two subcases (under the case M = L), viz., $L < \lceil NR \rceil$ and $\lceil NR \rceil \le L$. In the first subcase we may apply Lemma 2 to the numerator of the right hand side of equation (8), yielding:

$$C(R,N,L,L) = \frac{(4)}{\binom{N+1}{L+1} - \binom{\lceil NR \rceil}{L+1}} = \frac{\binom{\lceil NR \rceil}{L+1}}{\binom{\lceil NR \rceil}{L+1}} = \frac{\binom{\lceil NR \rceil}{L+1}}{\binom{N+1}{L+1}}$$

$$= 1 - \frac{\binom{\lceil NR \rceil}{L+1}}{\binom{N+1}{L+1}}$$
(9)

In the second subcase we invoke the fact, from the definition of binomial coefficients, that

$$K < L \implies \binom{K}{L} = 0 \tag{10}$$

This implies that the summands of the right hand side of equation (8) equal zero until I = L. Therefore

$$\lceil NR \rceil \le L \implies \sum_{I=\lceil NR \rceil}^{N} \binom{I}{L} = \sum_{I=L}^{N} \binom{I}{L} \implies$$

$$\implies \text{the numerator and the denominator of equation (8) are the same} \implies$$

$$\implies C(R,N,L,L) = 1$$

But $\lceil NR \rceil \le L \implies \lceil NR \rceil < L+1$, so if we apply implication (10) to the numerator of the rightmost term of equation (9) we get exactly the same result, viz., C(R,N,L,L) = 1 - 0 = 1. Therefore using implication (10) allows equation (9) to work for both subcases. Therefore we have proved:

$$C(R,N,L,L) \stackrel{\underline{(9)}}{=} 1 - \frac{\binom{\lceil NR \rceil}{L+1}}{\binom{N+1}{L+1}} = \frac{1 - \frac{\lceil NR \rceil!}{(L+1)!(\lceil NR \rceil - L - 1)!}}{\binom{(N+1)!}{(L+1)!(N-L)!}} = \frac{1 - \frac{\lceil NR \rceil!}{(L+1)!(N-L)!}}{\binom{(N+1)!}{(L+1)!(N-L)!}}$$

Page 9 of 19.

$$= C(R,N,L,L) = 1 - \frac{\lceil NR \rceil! (N-L)!}{(\lceil NR \rceil-L-1)! (N+1)!} =$$

$$= C(R,N,L,L) = 1 - \frac{I = \lceil NR \rceil-L}{N+1} =$$

$$= 1 - \frac{L}{I = 0} = \frac{\lceil NR \rceil-I}{(N+1)-I}$$
(11)

It should be noted that evaluating the right hand side of equation (11) requires little more than multiplying together L+1 simple (i.e., just integer-over-integer) fractions. Thus equation (11) can be handled easily on any hand calculator, and even more trivially on any programmable calculator (unlike equation (1), i.e., equation (4) in Reference [R1] or equation (1) in Reference [R2], which generally requires a computer).

8. Corollary 1:
$$\frac{N-1}{N} < R \implies C(R, N, L, L) = \frac{L+1}{N+1}$$
Proof:
$$\frac{N-1}{N} < R \implies N-1 < NR$$

But R \leq 1 = \triangleright NR \leq N . Therefore we have N-1 < NR \leq N . Therefore

$$\frac{N-1}{N} < R \implies \lceil NR \rceil = N \tag{13}$$

Therefore
$$\frac{N-1}{N} < R$$
 $\frac{(13)}{C(R,N,L,L)}$ $\frac{(11)}{1}$ $1 - \sum_{I=0}^{L} \frac{(N-L)+I}{(N-L+1)+I} =$

$$= 1 - \frac{(N-L)+O}{(N-L+1)+L} =$$

$$= 1 - \frac{N-L}{N+1} = \frac{L+1}{N+1} , \text{ q.e.d.}$$

Page 10 of 19.

An alternative proof of implication (12), which is direct in the sense that it does not use equation (11) (i.e., which treats implication (12) as a result in its own right rather than as a corollary) is offered as Appendix D.

9. For example, suppose the Air Force has a type of airplane of which there are only 9 copies, is interested in a reliability of R = .9, has tested 5, and all have passed. Then the conditions of Corollary 1 are satisfied, and the confidence level in a fleet reliability of 90% is, by implication (12),

$$C(.9,9,5,5) = \frac{L+1}{N+1} = \frac{5+1}{9+1} = 60\%$$

If the reliability of interest is R = .9 then of course Corollary 1 would apply to any fleet size less than 9 also. A real example of such a small population of interest might be the E-4B fleet. (Of course no homogeneity assumption is being invoked here. In fact, no more than the maximum ignorance assumption is being employed. -- For a treatment of this issue cf. Reference [R2].)

10. Returning to paragraph 2, above, we now seek an analytic simplification of equation (1) for M = 0 (i.e., all tests failed). Well,

$$C(R,N,L,0) \stackrel{(2)}{=} 1 - C(1-R,N,L,L-0) = \frac{(11)}{1 - \prod_{I=0}^{L} \frac{([N(1-R)]-L)+I}{(N-L+1)+I}} = \frac{L}{1-0} \frac{C(R,N,L,0)}{(N+1)-I} = \frac{L}{1-0} \frac{[N(1-R)]-I}{(N+1)-I}$$

$$= C(R,N,L,0) = \frac{L}{I-0} \frac{[N(1-R)]-I}{(N+1)-I}$$

$$= (14) .$$

As was the case with equation (11), evaluating the right hand side of equation (14) requires little more than multiplying together L+1 simple (i.e., just integer-over-integer) fractions; this is easily done on any hand calculator.

11. Corollary 2:
$$R < \frac{1}{N} \implies C(R, N, L, 0) = \frac{N-L}{N+1}$$
 (15).

Proof:

$$R < \frac{1}{N} = \frac{N - (N-1)}{N} = 1 - \frac{N-1}{N}$$

$$\sim$$
 $\frac{N-1}{N}$ < 1-R

Therefore

$$R < \frac{1}{N} \implies C(R,N,L,0) \xrightarrow{(2)} 1 - C(1-R,N,L,L) \xrightarrow{(12)}$$

$$\frac{(12)}{1 - \frac{L+1}{N+1}} = \frac{N-L}{N+1} , \text{ q.e.d.}$$

Of course an alternative way of proving implication (15) would be to deduce it from equation (14); hence the term "corollary".

- 12. Obviously, by changing Lemmas 1 and 2 appropriately results (14) and (15) could be derived first, and then results (11) and (12) obtained from them by application of the Symmetry Theorem for Confidence in Reliability of a Finite Population (equation (2)).
- 13. In closing, it might be helpful to make a few remarks concerning the relationship between these results and some other facts in the time-independent-reliability field. Suppose the cardinality of the population of elements each of which might be either a "success" or a "failure" is infinite. Then equation (10) in Reference [R4] plays the same role for such an infinite population as equation (11) in the present note plays for finite populations. Sometimes "success" is defined to be a physical variable's having a value in a certain interval (or domain). If the interval is the data range with one end point deleted (thus making the interval half open), then equation (13) in Reference [R4] is the appropriate equation if the population of possible realizations of that physical variable is infinite. However the present note provides no analogue for this latter case for finite populations since then M = L-1 . -- Very similar in form to equations (10) and (13) of Reference [R4] are two variations of Wilks's Tolerance Theorem. See for example equations (13.6.4) and (13.6.12) of Reference

[R5], or equation (22-10) of Reference [R6] (there is a sign error in at least one printing of the latter), or any of several references provided under "Collateral Reading" at the end of § 22-6 of Reference [R6]. The differences in form between these variations of Wilks's Tolerance Theorem and equations (10) and (13) of Reference [R4] can be attributed to differences between the technical definitions of "confidence" being used (viz., Bayesian vs Neyman-Pearson). The point, however, is that it would be incorrect to use either the equations of Reference [R4] or those referred to in References [R5] and [R6] above if the population of possible realizations is finite. The results of using such infinite population results would be approximately cornect if the population is quite large ("essentially infinite") compared to the sample size; but the errors in using them when the population size is small, in contrast, can be considerable. (That Wilks's Tolerance Theorem is not exact for finite populations can be seen in several ways without tracking through the details of its proof. One way is simply to notice that the result is not a function of the population size. Another is simply to notice that γ in the equations referred to in References [R5] and [R6] above is a continuous function of β in those equations. Since it is known a priori that the fraction of finite population objects included in a fixed interval can be one of only a finite number of values (viz., the N+1 values $\frac{0}{N}$, $\frac{1}{N}$, $\frac{2}{N}$, ..., $\frac{N}{N}$), confidence in reliability of a finite population must instead be a step function, with steps at those values.)

Appendix A

Proof of Lemma 1.

 $= \frac{(N+1)!}{L!(N-L+1)!} = \binom{N+1}{L}$

Appendix B

Alternative proofs of Lemma 2.

Lemma 2:

$$0 \le J \implies \sum_{I=N-J}^{N} {I \choose L} = {N+1 \choose L+1} - {N-J \choose L+1} \qquad (4)$$

Alternative proof a: Adding up J+1 successive applications of Lemma 1, equation (3), yields:

Alternative proof b: Using $L \le N-J \le N$, so that $\sum_{I=N-J}^{N} = \sum_{I=L}^{N} - \sum_{I=L}^{N-J-1}$,

Lemma 2 can also be viewed as an immediate consequence of Lemma 3 (i.e., of eq. (6)) ... provided, of course, Lemma 2 was not used to prove Lemma 3 (i.e., provided Lemma 3 is proved as in the text, not as in Appendix C, below).

Appendix C

Alternative proofs of Lemma 3.

Lemma 3.

$$\sum_{I=1}^{N} \begin{pmatrix} I \\ L \end{pmatrix} = \begin{pmatrix} N+1 \\ L+1 \end{pmatrix} \tag{6}$$

Alternative proof a:

$$\sum_{I=L}^{N} \begin{pmatrix} I \\ L \end{pmatrix} = \begin{pmatrix} L \\ L \end{pmatrix} + \sum_{I=L+1}^{N} \begin{pmatrix} I \\ L \end{pmatrix} =$$

$$= 1 + \sum_{I=L+1}^{N} \begin{pmatrix} I \\ L \end{pmatrix} \qquad (16) .$$

Now choose $J \ni N-J = L+1$, i.e.,

$$J \stackrel{\triangle}{=} N-L-1 \tag{17}$$

Then equation (16) becomes:

$$\begin{split} \sum_{I=L}^{N} \begin{pmatrix} I \\ L \end{pmatrix} &= 1 + \sum_{I=N-J}^{N} \begin{pmatrix} I \\ L \end{pmatrix} & \frac{(4)}{L} \\ & \frac{(4)}{L} + \begin{pmatrix} 1 \\ L+1 \end{pmatrix} - \begin{pmatrix} 1 \\ L+1 \end{pmatrix} & \frac{(17)}{L} \\ & \frac{(17)}{L} + \begin{pmatrix} 1 \\ L+1 \end{pmatrix} - \begin{pmatrix} 1 \\ L+1 \end{pmatrix} & - \begin{pmatrix} 1 \\ L+1 \end{pmatrix} & = \\ & 1 + \begin{pmatrix} 1 \\ L+1 \end{pmatrix} - \begin{pmatrix} 1 \\ L+1 \end{pmatrix} & = \\ & 1 + \begin{pmatrix} 1 \\ L+1 \end{pmatrix} - 1 & = \begin{pmatrix} 1 \\ L+1 \end{pmatrix} & \text{, q.e.d.} \end{split}$$

Alternative proof b: If Lemma 2 has already been proved (without employing Lemma 3, of course), then substitute L for N-J in the two places where that difference occurs in eq. (4). The subtracted term on the right hand side of eq. (4) will then vanish, by implication (10). What remains will be eq. (6), i.e., Lemma 3, q.e.d.

Appendix D

A direct proof of Corollary 1.

Corollary 1:
$$\binom{N-1}{N} < R \longrightarrow C(R,N,L,L) = \frac{L+1}{N+1}$$
 (12)

Direct proof: Applying implication (13) to equation (8) yields:

$$\frac{N-1}{N} < R \Rightarrow C(R,N,L,L) = \frac{\binom{N}{L}}{\binom{N+1}{L+1}} = \frac{\binom{N(N-1) \cdot \cdots \cdot (N-L+1)}{L(L-1) \cdot \cdots \cdot 2}}{\binom{(N+1)N \cdot \cdots \cdot ((N+1)-(L+1)+1)}{(L+1)L \cdot \cdots \cdot 2}} = \frac{L+1}{N+1}, q.e.d.$$

References

- [R1] Confidence and Reliability in a Finite Population, by Chris Ashley. EMP Probability and Statistics Note 1; 18 February 1971.
- [R2] Effects of Supplemental Information on Confidence and Reliability, by Chris Ashley. EMP Probability and Statistics Note 7; 22 September 1973.
- [R3] Handbook of Mathematical Functions, by Milton Abramowitz and Irene A. Stegun. National Bureau of Standards Applied Mathematics Series 55; June 1964.
- [R4] <u>Confidence and Reliability in an Infinite Population</u>, by Chris Ashley. EMP Probability and Statistics Note 2; 7 October 1971.
- [R5] Introductory Engineering Statistics, by Irwin Guttman and S.S. Wilks. John Wiley & Sons, Inc.; copyright 1965.
- [R6] Introduction to Statistical Inference, by Harold Freeman. Addison-Wesley Publishing Company, Inc.; copyright 1963.