Probability and Statistics Notes

Note 1

Confidence and Reliability in a Finite Population

by
Chris Ashley

18 February 1971

Air Force Weapons Laboratory

Abstract

An approach to the problem of quantifying confidence in conclusions drawn from results of experiments is presented, assuming only a finite number of elements in the set from which the samples are being taken and assuming dichotomy has been imposed. The approach is applied to a hypothetical fleet of 1000 Minuteman missiles as an example. Confidence in reliability of such a fleet is discussed. Some calculated values are presented.

Preface

Experiments are conducted in order to suggest and justify conclusions and predictions about the subject experimented upon. The greater the amount of supporting experimental data gathered, the higher the confidence one may reasonably have in conclusions drawn from that data. Thus, confidence in a conclusion is a monotone increasing function of the amount of experimental evidence supporting the conclusion. It is possible to quantify this confidence precisely.

In this paper it is shown by reasoning directly from basic principles how to quantify confidence in the case where the population has been dichomotized by the experimenter. *No further assumptions are made regarding the distribution of the population.*

The reasoning begins with intuitively acceptable statements about confidence and proceeds without recourse to devices, such as Bayes' rule, use of which would be second nature for the statistician but which are foreign to many engineers. It is hoped that this avoidance of specialized terminology will make quantification of confidence more accessible and acceptable to the conscientious worker who lacks familiarity with the concepts of probability and statistics.

Summary

It is assumed that the experimenter is confronted with a population of N elements (where N is finite), of which K are in one class and the remaining N-K are outside that class. The experimenter knows the value of N, but does not know and needs to estimate the value of K. He makes his estimate by testing a random sample of L of the elements and finding that, of these L in his test subset, exactly M are within the class. Thus, M/L is the fraction of the test subset which is within the class. The theory developed in this paper can be used to show that M/L is the fraction which the experimenter can be most confident is the fraction of the total population within the class. More practically, the theory provides expressions for precisely quantifying the confidence which the experimenter may have that K is $\lfloor(M*N/L)+.5\rfloor$, or that K is any other value between 0 and N of interest to the experimenter.

As a particular application, a hypothetical fleet of exactly 1000 Minuteman missiles, each of which can be ascertained by test to be either "defective" (unable to complete mission) or "non-defective" (able to complete mission), is used.

Finally, the concept of system reliability is discussed. The example of a system of 1000 Minuteman missiles is continued. The paper ends by providing an expression for calculating the confidence, in any value of reliability of interest to the experimenter, as a function of the amount of experimental data available.

1. Your friend Sam has two barrels which contain 100,000 marbles each, all identical except for color. You know that all the marbles in one of these two barrels are white, and that the other barrel is evenly divided between 50,000 white marbles and 50,000 black marbles, all stirred up well together. Sam puts one of these barrels in front of you, but he doesn't tell you which one. You reach in without looking and pull out a great big double handful of marbles. If this double handful of well stirred marbles, say 100 of them, was solid white, didn't contain a single black marble, then you'd be pretty convinced that you'd had to deal with the barrel that had nothing in it but white marbles.

2. Another example which is even more obvious is two barrels of sand, in one of which all the grains are white and in the other of which only half are white and the other half are black, but otherwise identical. If the mixed barrel is thoroughly stirred, so that any grain you choose blindfolded is truly random, and the barrel as a whole has a salt and pepper appearance, than a double handful which was perfectly white and showed not a single black grain would pretty much persuade you that you'd been offered the homogeneous barrel to choose from.

3. Now let's quantify. Recall the barrels of marbles. If you drew out only one marble and it was white, then you'd take that as a more mild indication that it came from the pure white barrel. The reason is that there are twice as many ways for you to pull a single white marble out of that barrel as out of the mixed barrel, so it would be twice as likely you'd draw a white marble from the white barrel. (Read that sentence again.) In fact, we say your confidence in having had the white barrel set in front of you was in this case precisely twice as great as your confidence that the mixed barrel had been. Numerically, you'd be $66\frac{2}{3}$ % confident that you'd had to deal with the white barrel.

4. Now consider Minuteman. Assume there are exactly 1000 missiles in the fleet, and you have no idea beforehand what the number of defective

missiles is before you begin to test. So you pretend that someone has 1001 different fleets of Minutemen (each fleet analogous to a barrel in the example above), and that in each of his fleets there is a different number defective, from 0 through 1000. And you pretend that he has put one of these fleets in front of you, so that the number defective in the fleet you have to deal with could equally likely be any number from 0 to 1000. You conduct a test of L sites and discover that M of these L are defective.

5. Your reasoning then goes as follows. How many ways could I have chosen L different sites with M defectives from the fleet with 0 defectives in it? How many ways from the fleet with 1 ? How many ways from that with 2 ? And so on. If from the fleet with i defectives you could have done this in I ways, and from the fleet with j defectives you could have done the same thing in J ways, and I was twice as great as J, then the level of your confidence that you'd in fact had to deal with the fleet with i defectives would be twice as great as the level of your confidence that you'd had to deal with the fleet with j defectives. (Read that sentence again.) To get your net confidence that the fleet you had to sample from had exactly i defectives in it, you'd add up all the ways you could have tested L and found M defective in each of all 1001 of the possible fleets, and you'd divide that sum into the number of ways in which that could have happened with the i-defective fleet.

6. The next question is, therefore, how do you calculate the number of ways of drawing L elements at random, to contain M defectives, out of a population with a total of N elements among which are K defectives? (That is, how do you get the I and the J of paragraph 5, above?) We know that there are $\binom{K}{M}$ (the symbol for a binomial coefficient) distinct subsets of M defectives available to be drawn if there are a total of K defectives in the fleet. Similarly there are $\binom{N-K}{L-M}$ distinct subsets of L-M nondefectives available to be drawn from the total of N-K nondefectives in the fleet. The product of these two numbers is the number of ways of doing both at once, i.e., of drawing M defectives from K and L-M nondefectives from N-K, for a total draw of $M+(L-M) = L$, since if each

selection of a site is indeed random then the event of drawing the M defectives is independent of the event of drawing the L-M nondefectives. So if there is a total of K defectives in the fleet of N missiles, then there are $\binom{K}{M}\binom{N-K}{L-M}$ ways of drawing L missiles from the fleet for test of which exactly M are defective.

7. Let us now fix N, L, and M, and let I(K) be the number of ways of drawing L, among which are exactly M defectives, if the fleet contains a total of K defectives. Then

$$I(K) = \binom{K}{M}\binom{N-K}{L-M} \qquad (1) \quad .$$

By the reasoning in paragraphs 3 and 5, above, if $I(K_1)$ is twice as great as $I(K_2)$, then your confidence level that the total number of defectives in the fleet was K_1 would be twice as great as your confidence level that the total number of defectives was K_2. From this we see that, in general, confidence level in some feature of a population, given certain test subset characteristics, is simply the probability, before drawing the subset for testing, that such a subset would be drawn from a population exhibiting the suspected feature if one were drawing the subset at random from among all subsets of that size in all possible populations of that kind. If we let $\bar{C}(K)$ symbolize the confidence level that K is the total number of defectives in the fleet of N, after testing L and finding exactly M of the L defective, then in general

$$\bar{C}(K) = \frac{I(K)}{\sum_{J=0}^{N} I(J)} \quad .$$

A little thought will show that $I(J) = 0$ for $J < M$, since there obviously have to be at least M defectives in the fleet (you found that many). Similarly $I(J) = 0$ for $J > N-(L-M)$, since there have to be at least L-M nondefectives in the fleet. So we might as well write at once

$$\bar{C}(K) = \frac{I(K)}{\sum_{J=M}^{N-L+M} I(J)} .$$

Substituting from equation (1), this becomes

$$\bar{C}(K) = \frac{\binom{K}{M}\binom{N-K}{L-M}}{\sum_{J=M}^{N-L+M}\left[\binom{J}{M}\binom{N-J}{L-M}\right]} \qquad (2) .$$

8. Now, you aren't generally interested in whether the number defective in the fleet is exactly some number K . You are rather more likely to be interested in whether *at least* K of the N missiles could be expected to perform their missions. If you can rely upon it that no more than a fraction $\frac{K}{N}$ of the fleet will fail to perform its mission, i.e., that at least a fraction $1 - \frac{K}{N}$ will perform its mission, then we say the fleet has a reliability R of $1 - \frac{K}{N}$, i.e.,

$$R = 1 - \frac{K}{N} .$$

Solving this equation for K, incidentally, yields

$$K = N(1-R) \qquad (3) .$$

The confidence level you have in the reliability R of the fleet, then, is the sum of your confidence levels for each possible number of defectives from 0 to K. If we let C(R) symbolize your confidence level that the fleet reliability is R, then

$$C(R) = \sum_{I=0}^{K} \bar{C}(I) .$$

Again, $\bar{C}(I) = 0$ for $J < M$, since there must be at least M defectives in the fleet, so we might as well write

$$C(R) = \sum_{I=M}^{K} \bar{C}(I) \quad .$$

By equation (3), this is

$$C(R) = \sum_{I=M}^{N(1-R)} \bar{C}(I) \quad .$$

By equation (2), this becomes

$$C(R) = \sum_{I=M}^{N(1-R)} \left\{ \frac{\binom{I}{M}\binom{N-I}{L-M}}{\sum_{J=M}^{N-L+M} \left[\binom{J}{M}\binom{N-J}{L-M}\right]} \right\} \quad .$$

Since the denominator is formally independent of I, we may rewrite this as

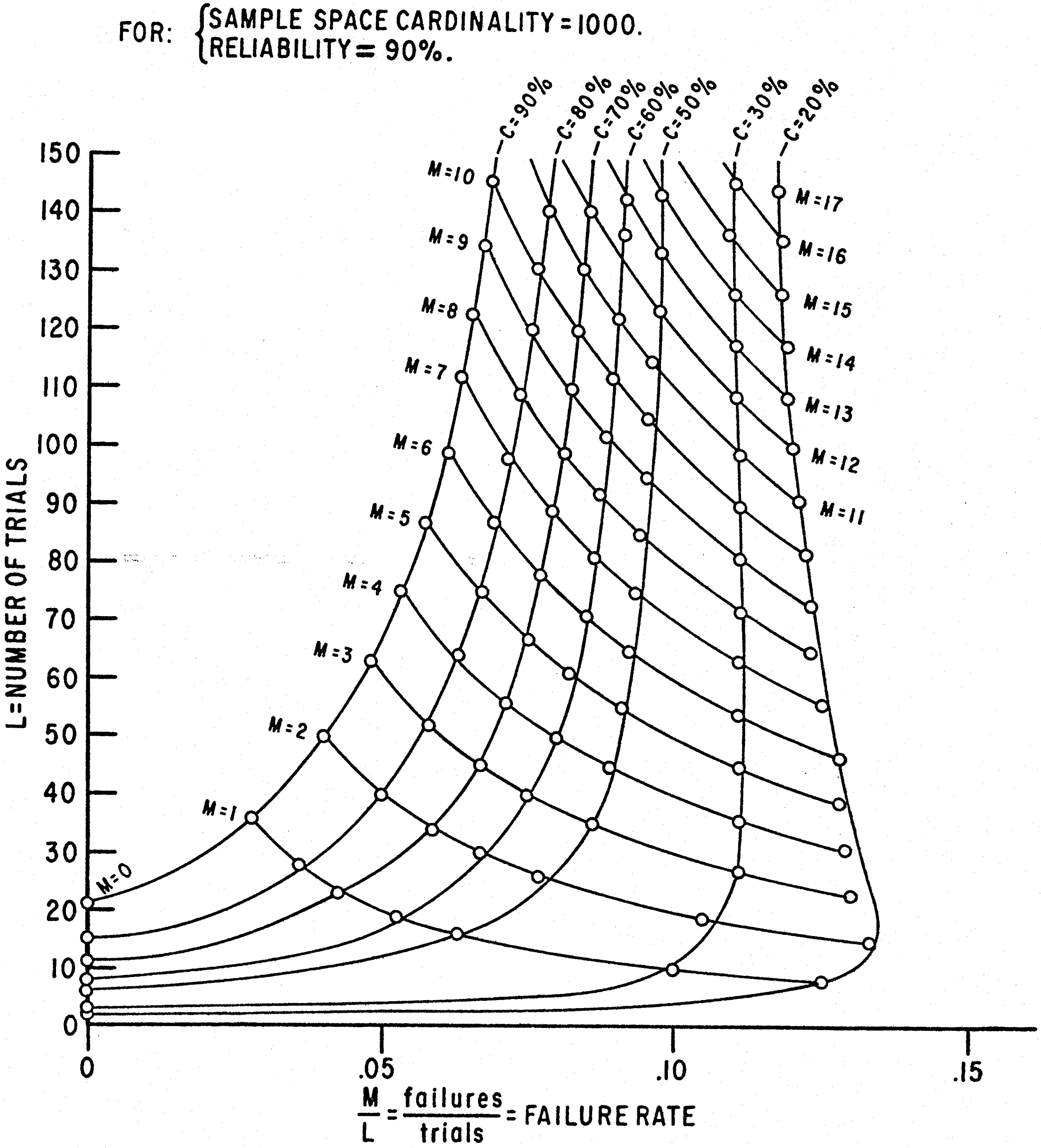
$$\boxed{C(R) = \frac{\sum_{I=M}^{N(1-R)} \left[\binom{I}{M}\binom{N-I}{L-M}\right]}{\sum_{J=M}^{N-L+M} \left[\binom{J}{M}\binom{N-J}{L-M}\right]}} \qquad (4) \quad .$$

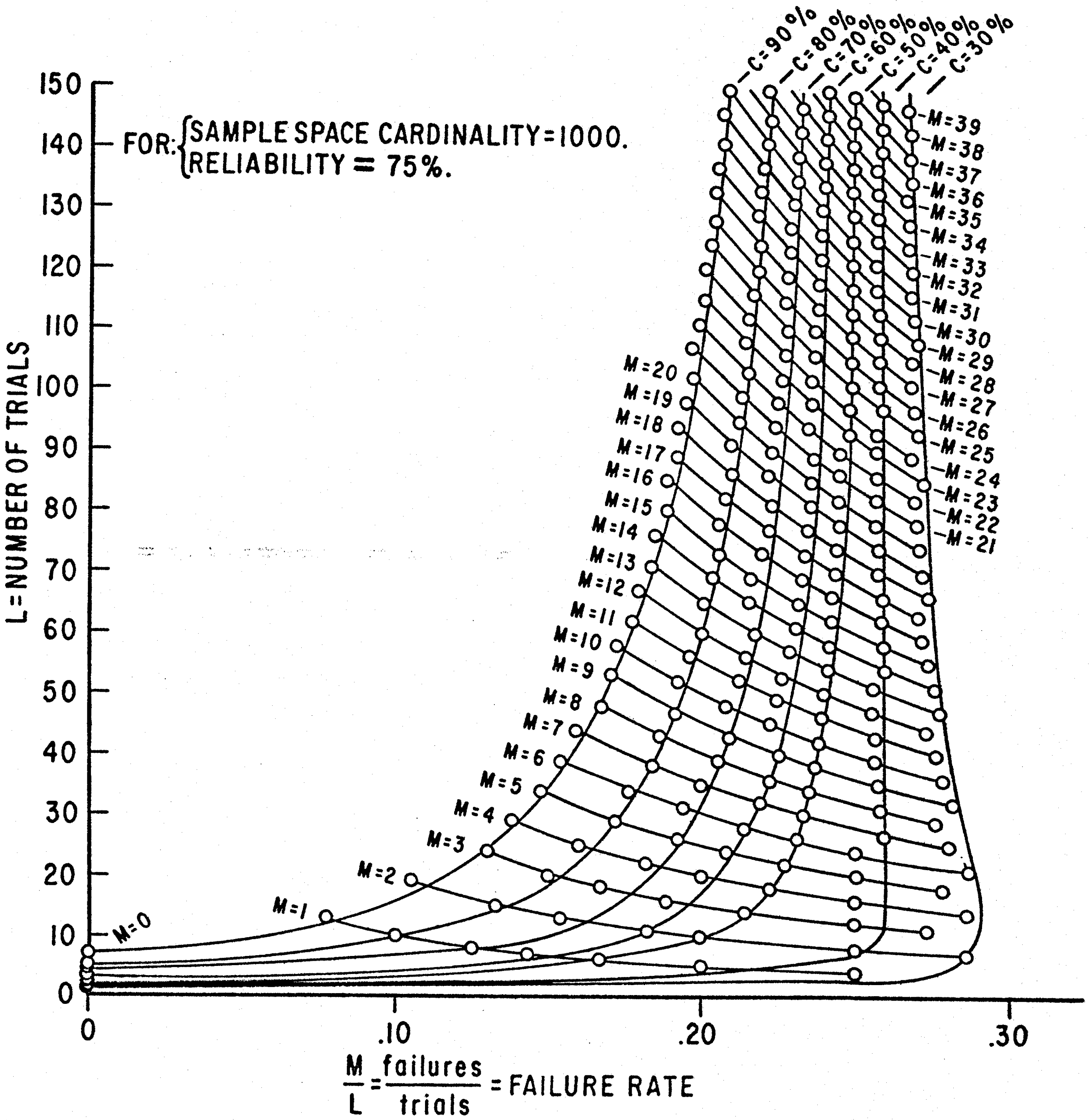
9. It would be optimistic to assume that you would in general be able to evaluate this expression with no more help than from a pencil and paper. So there now exist digital computer programs for evaluating it. For example, for a fleet size of 1000, if 5 missiles had been tested and it had been concluded from the tests that exactly 1 of these 5 would

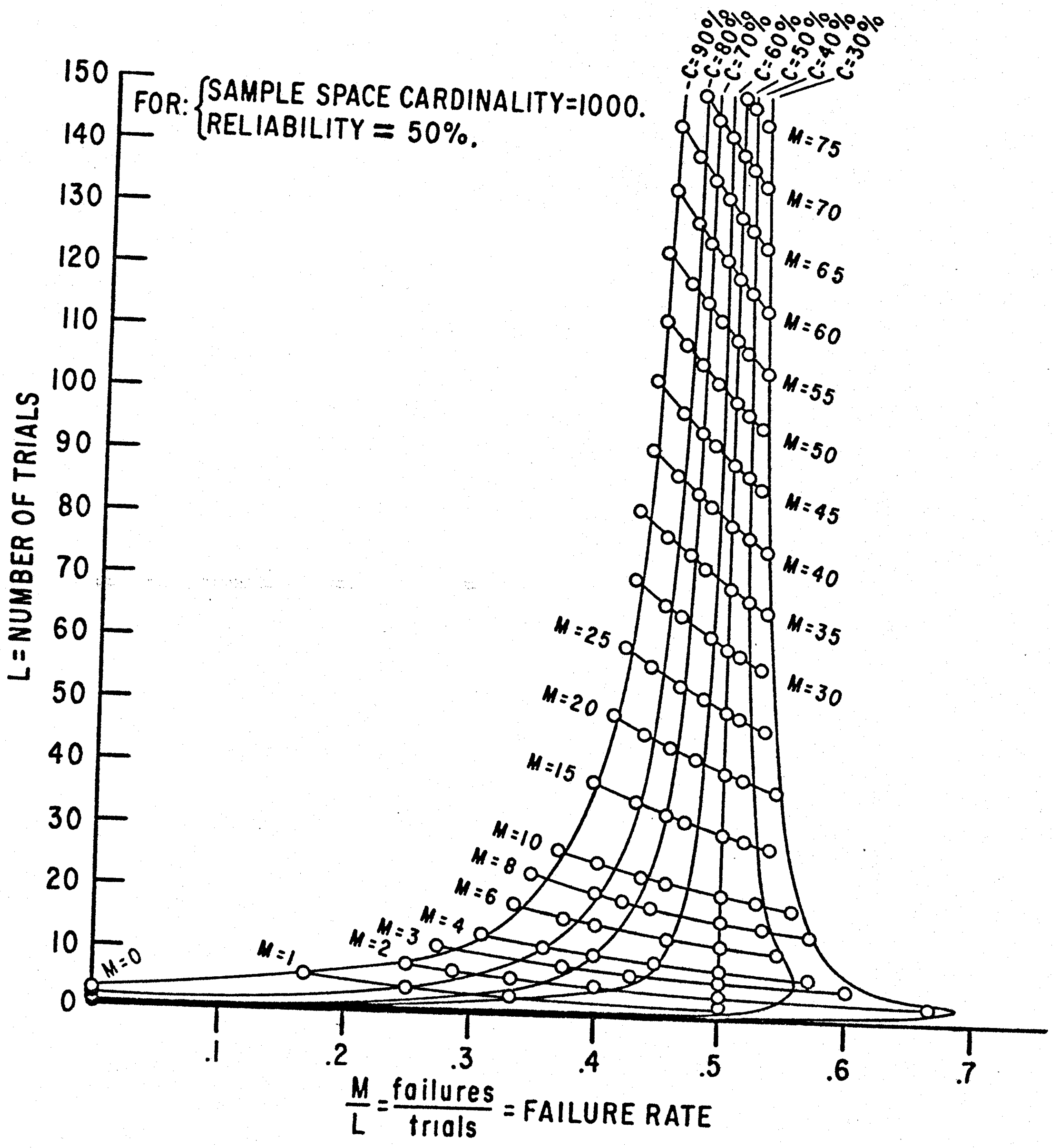
have failed to perform its mission, then we would have $N = 1000$, $L = 5$, and $M = 1$. If we were interested in reliability of 65%, then $R = .65$, and the computer would tell us that $C(R) = .68284$. That is, we would be justified in having a confidence level of 68.284% in fleet reliability of 65%. (The attached graphs give further values.)

10. Note that equation (4) is in five variables, viz., the population size N, the size L of the tested subset, the number M of those that failed the test, the reliability R, and the confidence level C(R). The confidence level C is offered as a function of the other four variables; we could write C(R,N,L,M). It is not always the case that R, N, L, and M are the independent variables in a particular application and C the dependent variable. In a particular application, for example, the problem may instead be, "I have a fleet of N missiles. If no more than one missile failed in test, how many would I have to test to achieve a confidence level of 95% in 90% fleet reliability?" In this case L is the dependent variable and N, M, R, and C are the independent variables. We would then need an algorithm for finding L(N,M,R,C). It may not be perfectly obvious at first glance how equation (4) is to be inverted to yield L(N,M,R,C). Fortunately, however, there already exist computer programs for solving such other permutations of the problem.

11. Finally, such formulas as equation (4) can be used in any problem where N is finite and dichotomy is possible. These are sometimes called hypergeometric problems. You could, for example, apply equation (4) to Sam's marbles (surely you remember Sam, from paragraph 1), or to a well mixed container of sand the top surface of which appeared entirely white, or to other aspects of Minuteman missiles than complete defectiveness.


FOR: SAMPLE SPACE CARDINALITY = 1000.
RELIABILITY = 90%.
C=90%
C=80%
C=70%
C=60%
C=50%
C=30%
C=20%
M=0
M=1
M=2
M=3
M=4
M=5
M=6
M=7
M=8
M=9
M=10
M=11
M=12
M=13
M=14
M=15
M=16
M=17
L=NUMBER OF TRIALS
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
0
.05
.10
.15
$\frac{M}{L} = \frac{\text{failures}}{\text{trials}}$ = FAILURE RATE


FOR: SAMPLE SPACE CARDINALITY=1000.
RELIABILITY = 75%.
C=90%
C=80%
C=70%
C=60%
C=50%
C=40%
C=30%
M=0
M=1
M=2
M=3
M=4
M=5
M=6
M=7
M=8
M=9
M=10
M=11
M=12
M=13
M=14
M=15
M=16
M=17
M=18
M=19
M=20
M=21
M=22
M=23
M=24
M=25
M=26
M=27
M=28
M=29
M=30
M=31
M=32
M=33
M=34
M=35
M=36
M=37
M=38
M=39
L=NUMBER OF TRIALS
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
.10
.20
.30
M/L = failures/trials = FAILURE RATE


FOR: SAMPLE SPACE CARDINALITY=1000.
RELIABILITY = 50%.
C=90%
C=80%
C=70%
C=60%
C=50%
C=40%
C=30%
M=75
M=70
M=65
M=60
M=55
M=50
M=45
M=40
M=35
M=30
M=25
M=20
M=15
M=10
M=8
M=6
M=4
M=3
M=2
M=1
M=0
L=NUMBER OF TRIALS
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
.1
.2
.3
.4
.5
.6
.7
M/L = failures/trials = FAILURE RATE