Quantifying Probability of Detection (POD) Using the Binomial Distribution

By Jennifer Brown and Chelsea Dalton

Q How can probabiity of detection (POD) be estimated for a given target size using find/no-find data from a nondestructive testing (NDT) system?

A In the aerospace industry, an NDT inspector may be required to find a crack-like defect of a given critical size with at least 90% POD and 95% confidence in order for certain hardware to meet design requirements. To demonstrate that NDT reliability meets these requirements, an inspector is tested to obtain data that can be analyzed statistically. When the inspection response is "find" or "no find," an estimate of POD for a given target size can be obtained by designing the test as a binomial experiment.

A binomial experiment consists of n independent Bernoulli trials. A Bernoulli trial is a single probabilistic experiment that results in either a "success" or "failure," where the probability of success is a fixed value p. What constitutes a success is defined by the problem and does not necessarily represent a desirable event. Though a crack of critical size in hardware is an undesirable event, for example, finding it with NDT is a "success."

The result from a single Bernoulli trial provides little information about the underlying process governing the trial. For example, the outcome (either find or no find) for one crack-like defect of critical size does not yield much information on POD. However, the collection of outcomes from repeated Bernoulli trials (e.g., find/no find data from several crack-like defects of critical size) does and is described by a binomial distribution with probability function

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (p)^{n-x} \quad for\ 0 \le x \le n$$

where P($X = x$) is the probability of x successes in n independent trials, p is the fixed probability of success on a single trial, and $\binom{n}{x} = \frac{n!}{x! \cdot (n-x)!}$ is the number of ways x successes could occur in n trials.[1]

The assumption of n independent trials with a fixed probability of success on a single trial is key and can be illustrated using a deck of cards. Suppose the binomial probability function will be used to calculate the probability of drawing one ace (a success) in five trials. In this example, p is known since the number of aces (4) and the total number of cards in a deck (52) are both known. However, the value of p depends on how the experiment is conducted. A binomial experiment requires that a card drawn from the deck be placed back into the deck and the deck shuffled prior to the next draw. When cards are drawn *with replacement,* the known value of p is 4/52 = 0.08 and is the same for each trial, maintaining the independence of each trial. If an ace is drawn on the first draw and not returned to the deck, then on the second draw, p is 3/51 = 0.06. The probability of success is no longer the same for the second draw because it is dependent on the first draw. Hence, designing the experiment such that trials are independent with the same probability of success is key to assuming a binomial experiment and using the binomial probability function.

In the NDT application, where x is the number of successful finds out of n independent inspection opportunities, it is not the probability of a given outcome P($X = x$) that is of interest but the value of p, which represents the POD for a single trial. Unlike with the deck of cards, p is unknown. However, the Law of Large Numbers says that x/n provides a good estimate since it approaches p as n increases.[2] If an NDT inspector, for example, is presented with 29 independent inspection opportunities to find a crack-like defect of a given target size and finds 28 out of 29, then an estimate of p = POD based on the sample data is 28/29 = 0.97. But what happens if the inspector finds all 29? Though p = 29/29 = 1 is an estimate, it is not a completely satisfactory one as in reality, the inspection cannot be perfect.

Because variation is always present and can never be entirely eliminated, a confidence bound is put around an estimate based on a single sample of data to take into account uncertainty due to sampling. When quantifying POD for a given target size, a one-sided lower confidence bound on p is calculated to estimate the *minimum* POD with 95% confidence using the Clopper-Pearson Method,

or "The Conservative Method," which assumes a binomial experiment.[3] When there are 29 successes out of 29 independent trials, the Clopper-Pearson Method results in a *minimum* POD estimate of 90% with 95% confidence. This is the basis for the point estimate method (PEM).

The PEM is one method used to qualify NDT inspectors in the aerospace industry to perform a find/no-find inspection for crack-like defects on hardware that needs a minimum 90% POD with 95% confidence for a given target size to meet design requirements.[4]

The PEM qualification test presents the inspector with 29 independent inspection opportunities to find a crack-like defect of a given target size. If all 29 are found, the design requirement is met. If one is missed, the inspector may be presented with 17 additional independent inspection opportunities. If all 17 are found, resulting in 45 total finds out of 46 total independent opportunities, then a minimum 90% POD with 95% confidence for the given target size can be claimed based on the Clopper-Pearson Method. (Table 1 provides a lower bound estimate of p with 95% confidence for other values of n.)

Number of "Failures" (x = 1 – Number of Failures)	0	1	2
Number of Independent Bernoulli Trials, *n*	Lower bound on *p* rounded down to the nearest hundredth		
10	0.74	0.60	0.49
20	0.86	0.78	0.71
30	0.90	0.85	0.80
40	0.92	0.88	0.85
50	0.94	0.90	0.87
60	0.95	0.92	0.89

Table 1
One-sided Lower Bound on Success Probability *P* with 95% Confidence Using the Clopper-Pearson Method

Though qualification of an NDT inspector using the PEM method is practically simple, the test design *must* meet the assumptions of a binomial experiment. Since POD is not the same for different defect types and is expected to increase with size for a given defect type, the 29 Bernoulli trials must represent 29 independent inspection opportunities that are representative of the same defect type and size in order for p = POD to meet the condition of being a fixed value on a single trial. Otherwise, the statistical integrity of the POD claim for a given target size that assumes a binomial experiment may be compromised. ∎

REFERENCES
1   Dunham, W. *The Mathematical Universe: An Alphabetic Journey Through the Great Proofs, Problems, and Personalities.* New York: John Wiley & Sons, 1994: 11-22.

2   Ibid.

3   Meeker, W.Q., Hahn, G. J., and Escobar, L.A. *Statistical Intervals: A Guide for Practitioners and Researchers.* Second Edition. New York: John Wiley & Sons, Inc., 2017: 103-104.

4   NASA Technical Standards System. "Nondestructive Evaluation Requirements for Fracture-Critical Metallic Components" (NASA-STD-5009).

Jennifer Brown is an independent statistical consultant with 15 years of experience in aerospace. She is a member of the committee on quality and statistics (E11) and the subcommittee on specialized NDT methods (E07.10).

Chelsea Dalton is an NDT Engineer at Blue Origin with a focus on NDT reliability assessments.

John Carson, Ph.D., is a senior statistician for Neptune and Co. and the Data Points column coordinator. He is a member of the committee on quality and statistics (E11), a former chair of the subcommittee on statistical quality control (E11.30), and a member of several other ASTM International committees.