

The Binomial Approach for Probability of Detection

Carlos Correia
Grupo Endalloy C.A. - Caracas - Venezuela
www.endalloy.net
carloscorreya@gmail.com

Abstract

The present work, presents three different approaches to calculate the Probability of Detection (POD) Using the Binomial Approach. The ASME BPVC Sec. V Art. 14 proposal is included. POD will be a major issue during near future inspection projects. The spread of fracture mechanics as first method to establish critical defect size asks for more accurate proved NDT procedures and inspection teams. New methodologies for maintenance are under development to include the POD as one of the relevant factors to be considered in Probabilistic Failure Assessment and Risk Based Inspection. The Binomial Approach presents the advantage to be an easy method but relevant drawbacks should be considered.

1. Introduction

During inspection, not all detectable discontinuities would be detected. The proportion of detected (detectable) discontinuities, could change from time to time, from one inspector to another inspector, between different teams of inspection or equipments or combinations. This fact is gaining importance over the NDT industry, and this has led to the study of the Probability of Detection, which refers to the probability to detect a detectable discontinuity (for easy of explanation we will incorrectly use the term "defect" instead discontinuity in some paragraph of this paper). It is important to keep in mind that POD calculations are made based on detectable indications of certain critical size. There is no sense to calculate POD for non detectable discontinuities ⁽¹⁾.

POD applied in Nondestructive Testing has started at NASA in the early on 1970s ⁽²⁾, while the phenomenon seems to be conceived or imported from other branches of technology as telecommunications, radar and medicine ⁽³⁾.

Most Nondestructive Testing (NDT) Methods and techniques are based on the interpretation of signals which comes from interactions of some physical stimulus with the medium and the discontinuity. The reception and interpretation of those signals could be affected by many different phenomena, like system noise or other unexpected signals coming from other processes. In that way, the inspection results could be affected.

NDT as any other technology is susceptible to errors. The consolidated techniques have well known drawbacks, and the human factor is a major issue: Fatigue, lack of proper training and experience. Improper procedures or devices shall also be included in the list of factors that can introduce error.

The importance of POD in the NDT industry could be seen through the increasing number of papers written in the last years as shown in **Figure 1** ⁽³⁾.

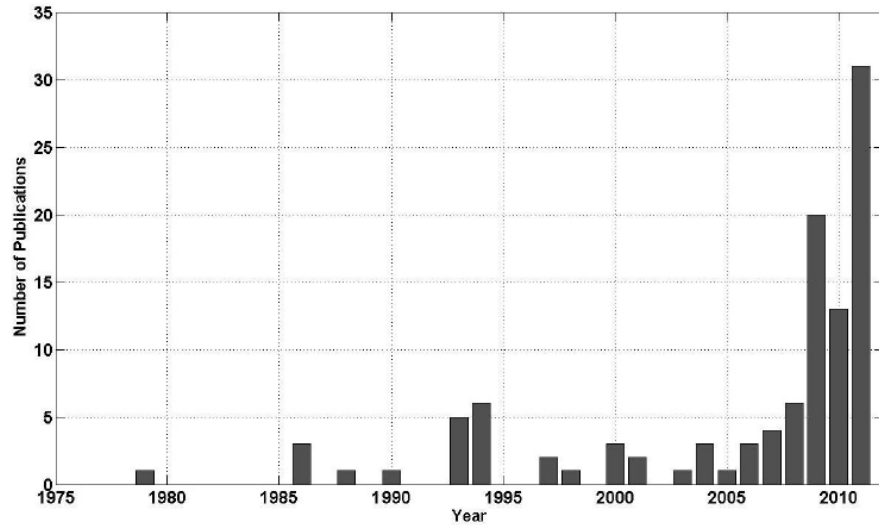


Figure 1. Number of publications on POD in the field of NDT.

Figure 1. (Courtesy of Jochen Kurz, Fraunhofer Institute for Non-Destructive Testing IZFP)⁽³⁾.

Some industry sectors as part of Nuclear, offshore oil and gas and Aeronautics consider POD as crucial, including the fact that POD will have an incidence in Risk⁽⁴⁾ and Probability of Failure⁽⁵⁾.

The present paper, shows some techniques and examples to have an estimate of POD from binomial point of view. It is recognized that the binomial approach is far from optimum method but simplicity of calculations make it an attractive choice. This is the technique indicated in ASME BPVC Section V, Art.14 (2013).

2. Foundations

If an experiment is performed with trials that are independent each other with only two possible outcomes (failure or success for example) with probability of success p , the phenomena can be described by a binomial distribution:

$$Bin(n, p) \quad (E.1)$$

where p is the actual proportion of success of the population.

The proportion of fails and success follow the rule:

$$p + q = 1 \quad (Ec. 2)$$

Where q is the proportion of fails.

Known p , the probability to have X success in n trials is:

$$p(x = X) = \binom{n}{X} p^X q^{(n-X)} \quad (Ec. 3)$$

But the unknown quantity is p , in the NDT case, the proportion of positive detections in our system, including all factors that could affect the test: environment, procedure, equipment, operators and the component under evaluation.

In the classical example the problem is described by cube plenty of green and red apples. If one need to know the proportion of red apples, we can take all the apples and count. This couldn't be done if the populations is too big. In this case, the method is take a sample of the population and calculate the sample proportion \hat{p} instead p .

If n trials where done with k successess, i.e. if a welded plate with $n = 10$ defects is examined and $X = 8$ defects where detected a rough estimate of success proportion (the POD) of the system is:

$$\hat{p} = \frac{X}{n} \quad (\text{E. 4})$$

$$\hat{p} = \frac{8}{10} = 0.8 \rightarrow 80\%$$

The first experiment seems to indicate that the POD could be 80%. If a new experiment is performed under the same conditions of previous one, it is possible that the "new" sample proportion, the new score will be 60%. If the experiment is repeated many times, is possible that different sample proportions will result each case. So, a conservative procedure is needed to fix a minimum POD from one experiment.

The Statistical way to deal with the previous situation is through the use of *confidence intervals*. In the present case, the confidence means that if the test is repeated a certain number of times, we have a percentage of guarantee (the confidence) that the parameter of interest is inside the interval; i.e. if the confidence is 95% and the calculated lower bound POD is 90%, we can be sure that in 95 of times the POD will be 90% or greater, as presented in **Figure 2**.

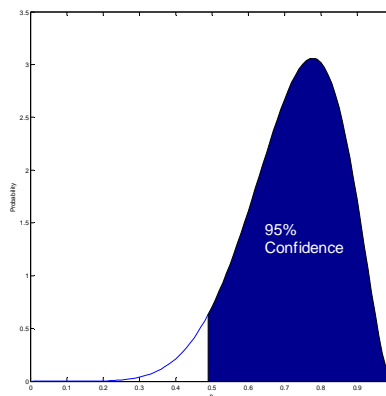


Figure 2. 95% Confidence upper Interval.

As p is unknown, it can be inferred with reasonable statistic support that ⁽⁵⁾:

$$L(\hat{p}) < p < U(\hat{p}) \quad (\text{Ec. 5})$$

In other words, if possible to create the interval from the test which produced \hat{p} . For this interval $L(\hat{p})$ is the lower bound and $U(\hat{p})$ is the upper bound of the interval.

After the interval has been built the lower bound $L(\hat{p})$ can be calculated and can be used as a good estimate of minimum possible POD for a confidence level.

Most times a POD of 90% is required with a confidence of 95%. The accepted nomenclature for this value is $POD_{90/95}$. This number seems to be selected from an arbitrary basis, nevertheless, some recent works connect the POD with failure probability and risk based inspection philosophy to produce specific requirements for POD^{(4) & (5)}.

In accordance with the previous discussion, for practical purpose the POD would be the lower bound $L(\hat{p})$ of the confidence interval:

$$L(\hat{p}) \leq p < 1 \quad (\text{E. 6})$$

3. Confidence interval

The confidence is equivalent to the area below the distribution function. If 95% confidence is required, this means that the interval should be defined by 95% the area below the curve ($\delta = 0.95$) as shown in **Figure 3**. Once the confidence has been fixed, the next step is to find the lower limit $k = L(\hat{p})$.

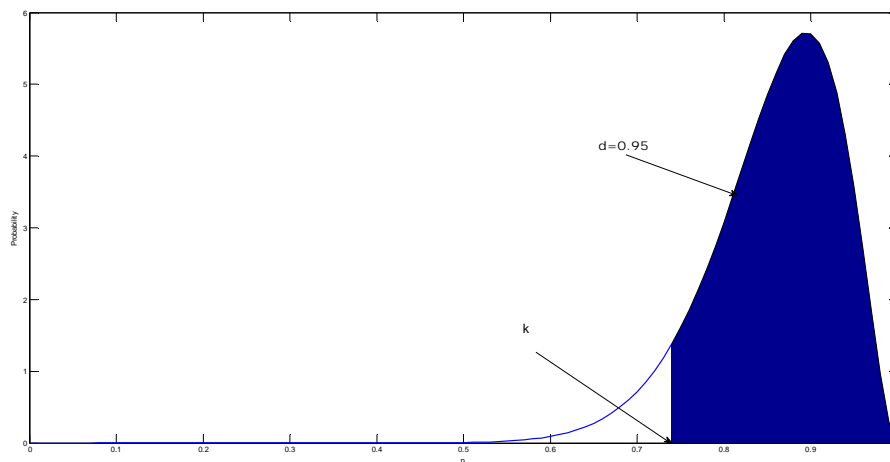


Figure 3. Lower bound confidence interval.

In **Figure 3**, an example of the *Beta Distribution Function* is presented. The lower bound limit k , has the following relationship with δ (the confidence coefficient) ⁽⁵⁾:

$$\delta = \int_k^1 \text{Beta}(p, a, b) dp \quad (\text{E.7})$$

Expression (E.7) could be used to calculate the confidence coefficient δ from $k = L(\hat{p})$ or vice versa (which would be the most common situation). Different software could be used. Annex A present a MATLAB script which let us find the confidence δ ($\text{Confidence} = \delta \times 100$) from $k = L(\hat{p})$.

4. Bayesian Approach

In the Bayesian approach, previous knowledge of an aspect related with the phenomenon, could be used to find a probability of occurrence; i.e. If we know the number of ice creams that was sold in a sunny Sunday, we can estimate if a specific day was a sunny Sunday from the number of ice creams sold that day.

In the Bayesian point of view, p is a random variable described by a probability distribution function, which is subjective and based on the experimenter "previous" knowledge. In that way, if the function is known, it is possible to fix a *Credible Set* (*the Confidence Interval concept applied in the Bayesian framework*) and find his lower bound limit using (E.7).

Once the experimenter select a previous distribution for p , the experiment is conducted and previous distribution is corrected taking in consideration the experiment results to produce a *posterior* distribution. In our case, posterior distribution could be the same type of function than previous one, but with new parameters. These can be written in the following way⁽⁵⁾:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta} \quad (\text{E. 8})$$

Where $\pi(\theta)$ is the previous distribution of the random variable θ , $\pi(\theta|x)$ is the posterior distribution under the new evidence x , and $f(x|\theta)$ is the sample distribution.

Appendix 1 of reference (5) presents a clear explanation of the Bayesian Approach, and the use of a Beta distribution to represent the random variable p .

The Beta distribution use the Beta function with parameters a and b which is also built from Gamma Function as shown below:

$$\text{Beta}(p, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} \quad (\text{E. 9})$$

Where Γ is the gamma function.

NOTE 1: In the present paper the notation $\text{Beta}(p, a, b)$ is preferred instead $\text{Beta}(x, a, b)$ or $\text{Beta}(a, b)$ to make emphasis in the fact that is a Beta distribution of the random variable p .

A relevant aspect is that Beta distribution is the conjugate of the Binomial distribution. In that way, the denominator of (E.9) can be resolved analytically⁽⁵⁾. The easy of calculation introduced by this fact is one of the factors that make The Beta distribution a preferred Distribution Function for p . Another relevant point is that Beta is defined in the interval $[0,1]$.

In the **Figure 4**, the $Beta(p, 100, 20)$ distribution function is presented.

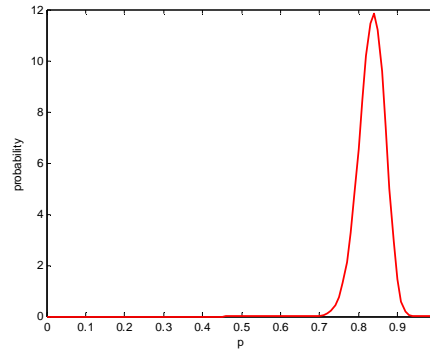


Figure 4. Beta($p, 100, 20$) Distribution function.

The Beta Distribution function shape is configured by the parameters a and b . If $a = 1$ and $b = 1$, the Beta distribution becomes the uniform distribution as can be observed in **Figure 5**.

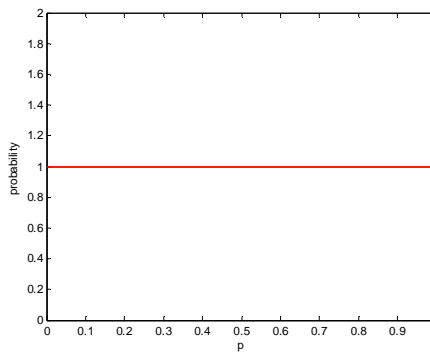


Figure 5. Uniform distribution, Beta($p, 1, 1$).

One of the most common approximations is to select Beta($p, 1, 1$), the uniform distribution as the previous distribution.

$$p_{previous} \rightarrow Beta(p, 1, 1) \quad (\text{Ec. } 10)$$

This is equivalent to say that the experimenter doesn't have any preference non previous knowledge of POD, all probabilities are equally possible.

A very appreciated property of Beta Distribution is that the new distribution parameters are easily obtained from test result following the next expressions:

$$a_{posterior} = a_{previous} + N_s \quad (\text{E. } 11)$$

$$b_{posterior} = b_{previous} + (N - N_s) \quad (\text{E. } 12)$$

Where N_s is the number of detected defects and N is the total number of defects included in the test.

The posterior distribution is:

$$p_{posterior} \rightarrow Beta(p, a_{posterior}, b_{posterior}) \quad (E. 13)$$

If $a_{previous} = 1$, $b_{previous} = 1$, (E. 14) becomes:

$$p_{posterior} \rightarrow Beta[p, (1 + N_s, 1 + (N - N_s))] \quad (E.14)$$

The new a parameter is the old a parameter plus the number of successes (detected defects). The new b parameter is the old b parameter plus the number of fails (non detected defects).

Some examples are presented below:

Example 1:

if we have a welded plate with 10 separated and similar defects (over critical size and with similar size each). We would like to calculate the POD of an inspection team using a well tested procedure and optimal equipment. After the inspection has been done, the team successfully detect 8 of 10 defects. So, $N = 10$ and $N_s = 8$.

We would like to find $L(\hat{p})$ the lower bound limit.

The posterior distribution in accordance with (E.14) after the test will be:

$$p_{posterior} \rightarrow Beta(p, 9, 3)$$

This distribution is shown in **Figure 6**.

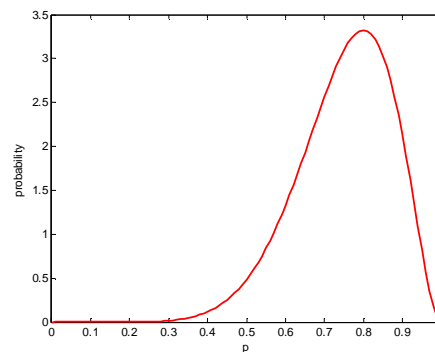


Figure 6. Beta(p,9,3) Distribution.

We need to fix the confidence required for the Test. Suppose 90% as a confidence required. The lower bound value is shown in **Figure 7**.

Under the Bayesian approach the *Credible Set* lower bound is $L(\hat{p}) = 0.58$ for 90% confidence. This means that if the test is repeated 100 times, 90% of the times $p \geq 0.58$. Under this conditions our POD could be represented by this value.

Although the score $\hat{p} = 0.8$ represent only one experiment and could not be used to evaluate the performance of the inspection team most part of times.

The lower bound value can be obtained through recognized PC statistics software even using free mobile devices applications ⁽⁶⁾.

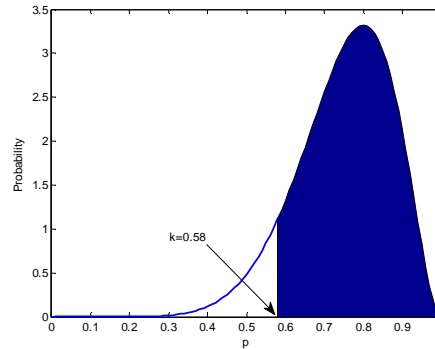


Figure 7. Lower bound limit for a confidence of 0.9.

Example 2

Suppose the same situation that Example 1 but the confidence required is now 95%. Posterior distribution is still: $p_{posterior} \rightarrow Beta(p, 9, 3)$

Figure 8 shows the *Credible Set* (Confidence Interval) and the lower bound limit.

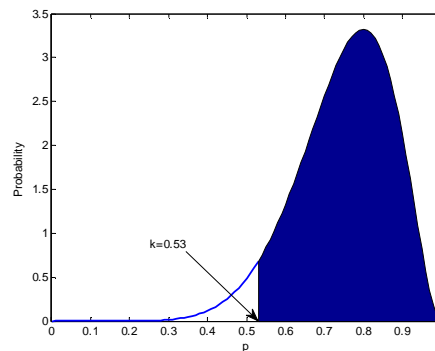


Figure 8. Lower bound limit for a confidence of 0.95.

Now: $L(\hat{p}) = 0.53$.

5 ASME BPVC Art. 14 (2013) POD Lower Bound

ASME code seems to use the Clooper Pearson approach also called the exact interval ⁽⁸⁾. Paragraph below ASME notation is used. The code presents the following expression referred to the lower bound confidence interval:

$$\alpha = \beta(p_{low}; D, N - D + 1) \quad (E.15)$$

The expression seems to be easily compressible in the following form:

$$\alpha = \int_{p_{low}}^1 Beta(p, D, (N - D) + 1) dp \quad (E.16)$$

Where p_{low} is the lower bound, D represent the successes (defects detected) and N is the total number of trials (the total number of defects) and α is the confidence coefficient.

The point of view of ASME is close as presented above (E.7). The main difference is the posterior parameter a , which is exactly equal to the number of successes $a_{posterior} = N_s$. The other parameter remains the same $b = (N - N_s) + 1$.

The posterior distribution after the test will be:

$$p_{posterior} \rightarrow Beta(p, N_s, (N - N_s) + 1) \quad (E.17)$$

The (E.17) produce the same Interval as the Clooper Pearson "Exact" interval as can be seen in the reference (8). More examples are presented below:

Example 3.

Suppose the same conditions as the Example 1 including the same confidence of 90%, $N = 10$ and $N_s = 8$. The posterior Beta distribution in accordance with (E.17) will be $Beta(p, 8, 3)$, as shown in **Figure 9**.

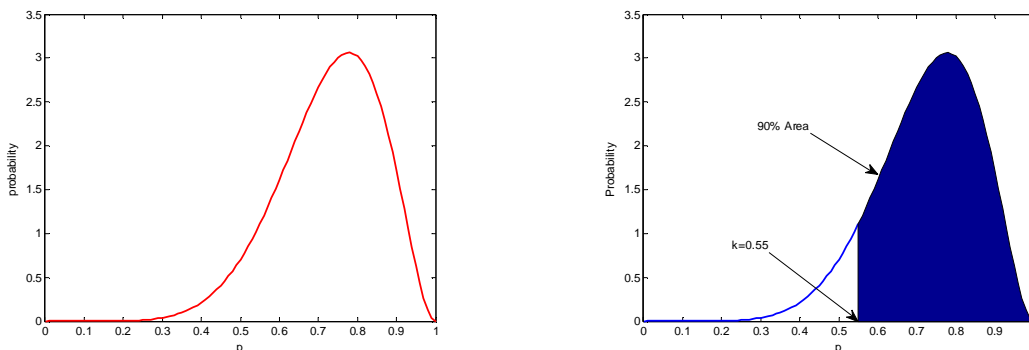


Figure 9. (left) Beta(p,8,3) Distribution Function. (right) 90% confidence interval and lower bound.

In this case $L(\hat{p}) = 0.55$ (as can be seen in one of the ASME examples).

Example 4.

Suppose the same conditions as Example 3, but the confidence is increased to 95%. In this case the new lower bound will be $L(\hat{p}) = 0.49$ as can be seen in **Figure 10**.

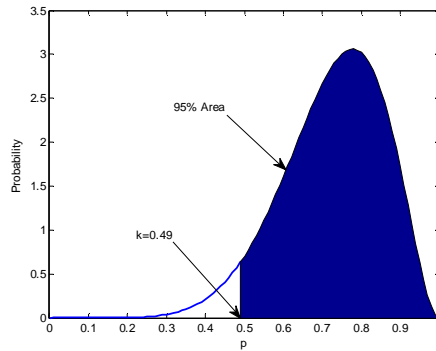


Figure 10. Beta(p,8,3) Distribution for 95% confidence interval and lower bound.

This results differs from example shown in ASME BPVC Sec V. Art. 14 (2013).

Example 5.

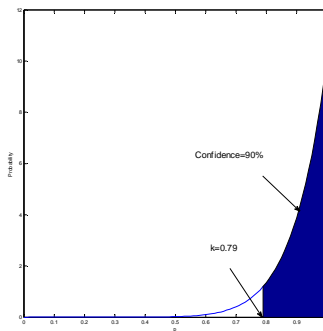
For the present case suppose the following values: Total number of defects in the test $N = 10$. Defects detected $N_s = 10$, the inspection team detects all the 10 defects. The lower bound is required for a confidence of 90%,95% and 99%.

According to (E.17) the distribution will be: $Beta(p, 10,1)$

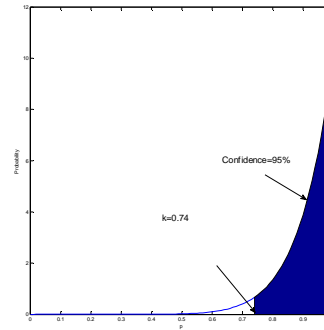
This Beta distribution produces the following results:

Confidence	Lower Bound
90%	0.79
95%	0.74
99%	0.63

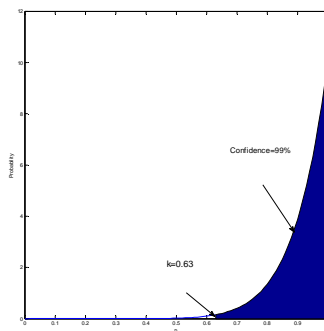
As shown in **Figure 11 (i), (ii) and (iii)**.



(i) Beta ($p, 10, 1$) Distribution with a 90% Confidence Interval.



(ii) Beta ($p, 10, 1$) Distribution with a 95% Confidence Interval.



(iii) Beta ($p, 10, 1$) Distribution with a 99% Confidence Interval.

Figure 11. Beta ($p, 10, 1$) Distribution with 90%, 95% and 99% Confidence Intervals.

The previous example shows a very relevant characteristic of the binomial approach and the reason why some references (including ASME and ENIQ documents) establish that this approach is far from optimum. As can be seen, even detecting all the defects included in the test, it is not possible to guarantee a high POD with a high confidence. If higher POD is requested the total number of defects used in the experiment should be increased.

If a $POD_{90/95}$ is required, the number of defects shall be increased to a minimum of 29, and the inspection team need to detect all 29 defects. This is known as the 29 of 29 rule for the binomial approach.

6 Some relevant aspects regard the Binomial approach presented in ASME BPVC Art. 14 (2013)

One of the drawbacks of the binomial approach is that all defects should have a very similar size and morphology. If the defects are of different sizes they has different probabilities to be detected and the foundations of binomial approach are not valid. Considerations about defect orientation and location could be also an issue.

If a "Size vs. POD " curve is needed, the test should be repeated for different groups of defects sizes. This increase amazingly the number of defects required. ASME gives the alternative to use the same plates with defects many times in a blind fashion, but this approach has a well supported criticism (see reference (1) MIL-HDBK-1823A (2009), page 36 point 6.4).

The most critical limitation of Binomial Test is the number of defects required to achieve the common $POD_{90/95}$. As Example 5 shows, 29 of 29 defects shall be detected to achieve $POD_{90/95}$. If the inspection team have only one miss a 90% POD could not be guaranteed and the lower bound will be $L(\hat{p}) = 0.84$ as can be seen in **Figure 12**.

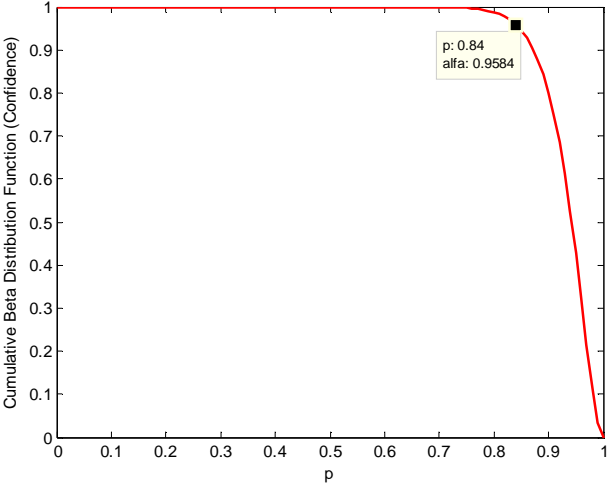


Figure 12. Confidence coefficients for Distribution Function Beta(p,28,2), 28 detected from 29 defects presented in the test.

If during the Test design, the experimenter assumes that maybe 2 defects would not be detected by the inspection team, 61 defects are needed to achieve $POD_{90/95}$ as can be seen in Table T-1471.1 de ASME BPVC Sec V. Art. 14 (2013).

One very important point, not frequently mentioned is that if the total number of defects is reduced, the performance of good inspection teams could be worse than if the number of defects is large.

7 Use of F-Distribution

One simple and fast approach is the use of the lower bound interval (*LCL*) mentioned in the reference (9). In this case the lower bound could be obtained using:

$$LCL = \frac{N_s}{N_s + (N - N_s + 1)F_\alpha(f_1, f_2)} \quad (E.19)$$

Where $F_\alpha(f_1, f_2)$ comes from the F-Distribution with coefficient of confidence α and f_1 y f_2 the distribution parameters, calculated using:

$$f_1 = 2(N - N_s) \quad (E.20)$$

$$f_2 = 2N_s$$

8 Conclusions

The Binomial approach is an easy way to calculate the POD but has some relevant drawbacks. If the confidence and POD should be high, the number of samples should be high with direct impact in the test cost. If the a Size vs. POD curve should be required, the test should be repeated for different groups of defects sizes.

9 Acknowledgments

Specially to Ed Ginzler (Materials Research Institute, Canada) for sharing the Clopper Pearson Spread Sheet file and his constant and dedicated support for all questions. Claudia Antonini (Universidad Simón Bolívar, Venezuela) for his support regard the Beta Binomial approach, Jochen Kurz (Fraunhofer Institute for Non-Destructive Testing IZFP, Germany) for share **Figure 1** and Matthew Bogner (University of Iowa, USA) for the mobile applet and the online Probability Distribution application .

Appendix 1. Confidence Associated with a lower bound value.

```
% Confidence calculation given the lower Bound
close all; clear all; clc;
x=0:.01:1;
e=8;% Successes
f=2;% Fails
alfa=e; beta=f+1;
%Beta Distribution Graph
y=betapdf(x,alfa,beta);
figure(1);
plot(x,y,'Color','r','LineWidth',2);xlabel('p');ylabel('probability')
plow=.49 %POD proposed Lower bound
Confidence=1-betacdf(plow,alfa,beta) %Confidence associated with 'plow'
% Carlos Correia - Jan 2015
```

9 Bibliography

1. MIL-HDBK-1823A. NONDESTRUCTIVE EVALUATION SYSTEM RELIABILITY ASSESMENT. <http://www.statisticalengineering.com/mh1823/index.html>
2. Ginzel E., Introduction to Statistics, <http://www.ndt.net/article/v11n05/ginzel2/ginzel2.htm>
3. Jochen H. KURZ.1, Anne JÜNGERT.2, Sandra DUGAN.2, Gerd DOBMANN. Probability of Detection (POD) determination using ultrasound phased array for considering NDT in probabilistic damage assessments. 18th World Conference on Nondestructive Testing, 16-20 April 2012, Durban, South Africa.
4. O. Cronvall, K. Simola et al. *A study on the effect of flaw detection probability assumptions on risk reduction achieved by non-destructive inspection*, Reliability Engineering and System Safety, 105 (2012), 90-96.
5. J. Garza, H. Millwater, *Sensitivity of probability-of-failure estimates with respect to probability of detection curve parameters*, International Journal of Pressure Vessels and Piping, 92 (2012) 84-95.
6. L. Gandossi, K. Simola, A BAYESIAN FRAMEWORK FOR THE QUANTITATIVE MODELLING OF THE ENIQ METHODOLOGY FOR QUALIFICATION OF NON-DESTRUCTIVE TESTING. Appendix 1. Institute for Energy, EUR 22675, 2007.
7. <https://play.google.com/store/apps/details?id=com.mbognar.probdist>. Matthew Bognar.
8. THULIN M., The cost of using exact confidence intervals for a binomial proportion. arXiv: 1303.1288v1. Mar 2013.
9. Generazio E., DIRECTED DESIGN OF EXPERIMENTS FOR VALIDATING PROBABILITY OF DETECTION CAPABILITY OF NDE SYSTEMS (DOEPOD). NASA, Hampton, VA 23681.