Journal of Modern Applied Statistical Methods

Volume 15 | Issue 2 Article 33

11-1-2016

Developing Bayesian-based Confidence Bounds for Non-identically Distributed Observations using the Lyapunov Condition

Garry M. Jacyna
The MITRE Corporation, gjacyna@mitre.org

Scott L. Rosen
The MITRE Corporation, srosen@mitre.org

Journal of Modern Applied Statistical Methods
November 2016, Vol. 15, No. 2, 536-562.
doi: 10.22237/jmasm/1478003460

ISSN 1538 − 9472

Developing Bayesian-based Confidence Bounds for Non-identically Distributed Observations using the Lyapunov Condition

Garry M. Jacyna
The MITRE Corporation

Scott L. Rosen
The MITRE Corporation

The purpose of this paper is to establish a direct method for assessing the confidence in the detection and identification probabilities for segmented observations that are not identically distributed across assigned segments within a region. This paper arrives at easily computable confidence intervals by showing through mathematical analysis that:

I. The probability of successful detection within each test segment can be characterized by a Beta distribution;
II. The distribution of a weighted sum of independent but non-identically distributed sample means is asymptotically Normally distributed by the Lyapunov variant of the Central Limit Theorem, i.e., the approximation improves as the number of samples increases;
III. Given that the distribution of the sample means convergences to a Normal distribution, the confidence intervals about the observed sample means for both the detection and identification probabilities can be determined in closed form for multiple target types.

The motivation for this approach is the need to determine the exceedance probabilities to support a Systems Acceptance Test based on collected data.

Keywords: Bayesian inference, analysis of designed experiments, beta distribution, Lyapunov condition

Background

A System Acceptance Test (SAT) requiring confirmatory data analysis (Box, Hunter, & Hunter, 2005) driven by apriori and politically deducted hypotheses is needed to assess the impact of a specific acquisition on two key system level performance parameters for a particular region: probability of detection (Pd) and probability of identification (Pid). The difficulty with this assessment is that

Email Garry Jacyna at: gjacyna@mitre.org. Email Scott Rosen at: srosen@mitre.org.

within this region of interest, only a finite portion is covered by each sensor's area-of-regard (AOR). In addition, there are an infinite number of threat compositions and avenues of ingress and egress (i.e., routes) that are possible throughout the area. Only a small sampling of operationally valid (traversable by the threat) routes across the region is executed and these are used to characterize the performance measures across the entire area. See Figure 1 for an illustrative view of this concept.

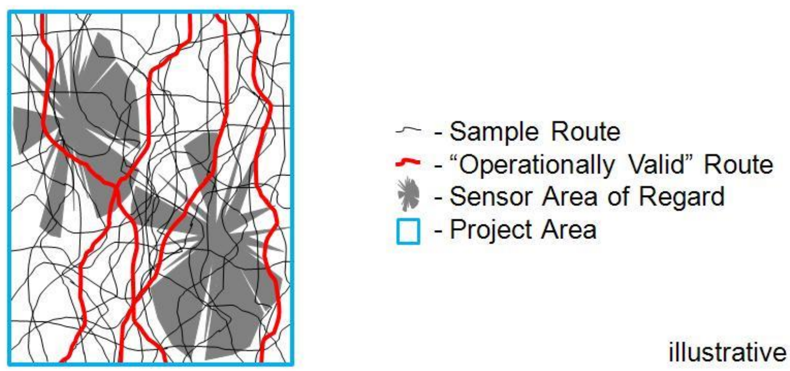


Figure 1. A high level illustration depicting the relationship among a project area and the sample routes, operational valid routes, and sensor areas-of-regards within it.

The overwhelming size of the test area introduces an additional test constraint. This is addressed by approximating the route samples with segment-level performance observations by considering a segment to be a contiguous subset of a given route. Moreover, a trial in this system acceptance test is defined as being a test observation made on a single segment. The subset of segments chosen for the test fall within a given sensor's AOR and belong to an operationally valid route as shown in Figure 2.

A difficult analysis problem arises when attempting to compute a system level estimate of performance involving an associated confidence bound and exceedance probability from segment-level observations made on small sample routes. This is because each route segment has a different underlining probability distribution that is a function of the different target/system/environmental factors present at the time of observation. Computing confidence intervals for Pd and Pid for individual segments is straightforward, but determining a single overall confidence interval for the sample set as a whole is not trivial since these independent random variables are drawn from different underlying detection and identification probability distributions. In this case it is not immediately apparent that the Central Limit Theorem (CLT) applies (Karr, 1993, p. 190-192).

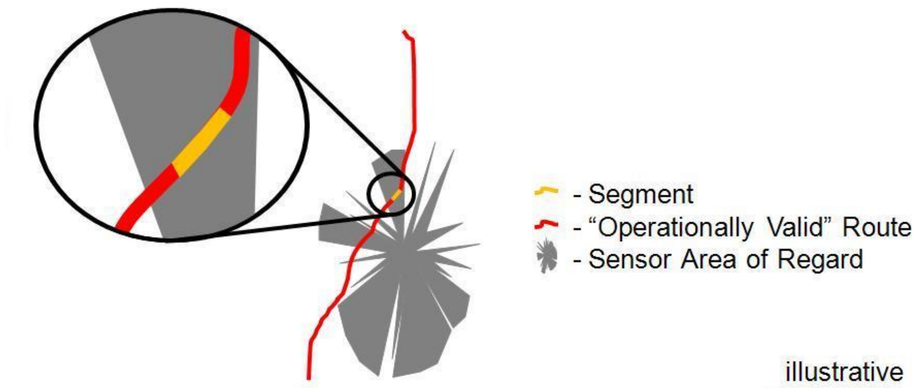


Figure 2. An illustration depicting the relationship among a segment, an "operationally valid" route, and a sensor AOR.

It is shown mathematically that the Lyapunov variant (Karr, 1993, p. 190-192) of the Central Limit Theorem (CLT) can be used to establish the Normality of the weighted sample means for Pd and Pid generated from this systems acceptance test. Furthermore, it is shown that a sample size of at least 1 trial for 10 unique segments is sufficient for approximating the resulting mean Pd and Pid observations by a Gaussian or Normal distribution.

Following this Normality result, the corresponding confidence intervals are then generated from the sample detection and identification proportions obtained in the test. Example computations are used to illustrate their implementation. A confidence interval calculator is then discussed for generating hypothetical confidence interval and exceedance probability values based upon inputted sample sizes for each individual segment and projected sample means for Pd and Pid. This calculator was then in turn used for shaping the experimental design for the test.

Introduction to the Analysis Problem

The work presented below describes, demonstrates, and justifies mathematically an approach for computing the confidence intervals associated with system level Pd and Pid observations. A few problem assumptions are necessary. These assumptions are as follows:

Problem Assumptions:

A. A detection trial pertains to the traversal of a single item of interest across an entire segment.

B. There is a binary outcome for an identification trial; the detection is successfully or correctly identified or the detection is unsuccessfully or incorrectly identified.

C. The sample probability of detection obtained from test constitutes the number of successful detections divided by the number of detection trials.

D. The sample probability of identification obtained from test constitutes the number of successful identifications divided by the number of identification trials

E. A single success probability p can characterize the probability of successful detection along a whole segment

F. A single success probability p cannot characterize the probability of successful identification across a whole segment, but can characterize the probability of successful identification for an individual identification trial within a segment.

G. The success probabilities for detection and successful identification are uniformly distributed between 0 and 1 across the sample set of segments.

Outline of the Approach

The probability density function (pdf) for the system Pd sample mean is derived as a function of segment-level observations from test. The analysis shows, through a mathematical proof and supporting Monte Carlo computations, that the distribution can be approximated as a Normal distribution.

In addition, through the application of a mixture distribution (across the identification trials within a segment), it is shown that the Normality results for Pd will also apply to Pid. Recommended confidence intervals then follow for Pd and Pid that are supported by this Normality result. This result is also valid in the case of a single target type or multiple target types as explained in the section focusing on Pid. A case study follows to illustrate how the reader can apply these confidence intervals. Below is a summary of the concepts that are presented and verified through mathematical analysis:

i) Through Bayes theorem, the probability of successful detection on each segment can be characterized by a Beta distribution.

   a. The weighted system detection probability is a convolution of Beta distributions and is referred to as an Augmented Beta Distribution.

   b. The distribution of the system sample mean is equivalent to the weighted system Pd distribution and, therefore, can also be characterized by the derived Augmented Beta Distribution.

ii) The Augmented Beta distribution is shown through a mathematical proof to be approximately normally-distributed by the Lyapunov variant of the Central Limit Theorem.

   a. The Lyapunov Central Limit Theorem specifies certain conditions that are sufficient to establish that the sum or average of a large number of independent observations is normally-distributed even if the observations are generated from different underlying probability distributions.

   b. The Lyapunov conditions hold when the threat arrival weights are uniformly distributed or when the selection of segments is equally likely.

   c. The Lyapunov conditions hold when the threat arrival weights are greater than zero for all but a finite number of segments.

   d. It is illustrated through an empirical computational study that the system sample mean rapidly converges to a Normal distribution within a 30 segment Test design alternative.

iii) Based on the fact that the system weighted sample means are approximately Normally distributed, the confidence about the observed sample means for Pd is:

$$C = \left(2\pi\sigma_\ell^2\right)^{-1/2} \int_{p^*}^{1} \exp\left\{-\left(s - m_\ell\right)^2 / 2\sigma_\ell^2\right\} ds$$
$$= F\left[\left(1 - m_\ell\right)/\sigma_\ell\right] - F\left[\left(p^* - m_\ell\right)/\sigma_\ell\right]$$

where: p^* is the exceedance probability or specified acceptable value for Pd, $F(x)$ is the Standard Normal Distribution:

$$F\left(x\right) = \left(2\pi\right)^{-1/2} \int_{-\infty}^{x} \exp\left(-s^2/2\right) ds,$$

m_ℓ and σ_ℓ^2 are the sample mean and variance, respectively:

$$m_\ell = \sum_{i=1}^{M} w_i \left(n_i + 1\right) / \left(N_i + 2\right), \quad \sigma_\ell^2 = \sum_{i=1}^{M} w_i^2 \frac{\left(n_i + 1\right)\left(N_i - n_i + 1\right)}{\left(N_i + 2\right)^2 \left(N_i + 3\right)}$$

and N_i and n_i are the total number of detection attempts and actual detections/identifications (see iv) observed in test, respectively, for each segment i, where M denotes the total number of segments.

iv) The Normality results for the weighted system sample mean also hold for the system Pid sample mean when the probability of successful identification on a segment is considered to be a mixture distribution as shown in the section focusing on Pid. The confidence interval can then be computed analogously as above. Furthermore, the Normality results also hold for multiple target types if the probability of detection is similarly considered to be a mixture distribution.

The Probability Density Function for Pd

This section develops the probability density function (pdf) for the weighted detection probability (Pd) across all segments assuming that the measurements relative to each segment are independent but not identically distributed. Each segment i has an associated test outcome $S_i = \{n_i, N_i\}$ which is a record of the number of detection successes n_i out of N_i possible trials. It is assumed that the outcome of each segment is statistically independent of any other segment outcome and that the unknown success probability for each segment is p_i. This implies that each segment n_i is binomially-distributed with known number of trials N_i and unknown probability p_i. The pdfs associated with each p_i are shown to be the well-known Beta distribution for an uninformative prior (i.e., a prior pdf that is uniformly distributed in the interval [0, 1]). Following that, this result is generalized for the weighted probability function $\sum_i w_i p_i$ for the system routes/segments when the segment outcomes S_i are independent but not identically distributed and the known arrival weights for each segment are given by w_i. A confidence interval is also determined for the probability that the weighted mean probability is greater than or equal to some exceedance probability p^*. The appropriateness of the Gaussian approximation to this general problem as the number of components in the weighted mean (i.e., number of segments) becomes large is then shown. Following that, an illustration of the approach is given to show that convergence to a Gaussian distribution is reached within the number of segments and trials allocated for test. Then a discussion of how the Normality result can be extend to Pid is provided.

The Beta Distribution for Segment Pd

Recognizing that the probability of detection p_i for a specific segment i is an *unknown parameter*, it is desirable to quantify this parameter with its own probability distribution. Now, determine the pdf associated with the probability of a successful detection p_i for a generic segment based on a fixed number of trials N_i and the number of successes from these trials n_i. From Bayes theorem (Bernardo & Smith, 2000, p. 241-255), this *conditional* pdf $f\,(p_i \mid n_i : N_i)$ is:

$$f_p\left(p_i \middle| n_i : N_i\right) = \kappa B\left(n_i, N_i; p_i\right) f_p\left(p_i\right), \tag{1}$$

where the corresponding likelihood function $B\left(n_i, N_i; p_i\right) = \binom{N_i}{n_i} p_i^{n_i} \left(1 - p_i\right)^{N_i - n_i}$ is the Binomial distribution, the prior distribution $f_p(p_i) = 1$ if $0 \le p_i \le 1$; otherwise, $f_p(p_i) = 0$ (using a uniformed prior assumption), κ is a proportionality constant and n_i denotes the number of successful detections out of N_i trials. It follows that $\kappa^{-1} = \int_0^1 B(n_i, N_i; p_i)\, f_p(p_i)\, dp_i$. Using the integral identity (Abramowitz & Stegun, 1965, p.258):

$$\beta(n+1, N-n+1) = \int_0^1 p^n (1-p)^{N-n}\, dp = n!(N-n)!/(N+1)!, \qquad (2)$$

where $\beta(x, y) = \int_0^1 p^{x-1}(1-p)^{y-1}$ is the Beta function, (1) can be rewritten as:

$$f_p(p_i | n_i; N_i) = p_i^{n_i}(1-p_i)^{N_i-n_i} / \beta(n_i+1, N_i-n_i+1). \qquad (3)$$

This density is referred to as the *Beta Distribution*.

A confidence interval is defined by C and p^*; C is interpreted as the probability that the true value of the unknown parameter p lies between the exceedance probability p^* and 1. In particular:

$$C = \text{Prob}(p_i \geq p^* | n_i, N_i) = \int_{p^*}^1 p_i^{n_i}(1-p_i)^{N_i-n_i}\, dp_i / \beta(n_i+1, N_i-n_i+1). \qquad (4)$$

The Augmented Beta Distribution Result for a System Pd Sample Mean

Suppose there are M segments with M associated test outcomes $S_1, S_2, \ldots, S_M$, where, as before, each set S_i records the number of trials, N_i, and the number of successful detections, n_i. It is assumed that the test results n_i are independent but *not* necessarily identically distributed. The unknown detection probability parameters associated with the M segments are labeled as $p_1, p_2, \ldots p_M$. The detection probability of the regional system should be represented by a weighted average of the segment detection probabilities, in which the segment weights w_i are computed from the relative proportion of threat traffic through the region expected to occur in segment i; $0 \leq w_i \leq 1$ and $\sum_i w_i = 1$. Therefore, the weighted average is a convex combination of the segment statistics $\ell = \sum_i w_i p_1$ and $0 \leq \ell \leq 1$. The value ℓ is a system-wide metric of detection performance.

To understand the regional system sample mean, examine the following joint pdf for the posterior detection probabilities across M segments. Using vector notation to express segment detection probability parameters, segment sample sizes, and number of segment detections as $\mathbf{p} = [p_1, p_2, \ldots, p_M]$, $\mathbf{N} = [N_1, N_2, \ldots, N_M]$ and $\mathbf{n} = [n_1, n_2, \ldots, n_M]$, respectively, this joint posterior probability is f_p (**p**| **n**; **N**). Because the sets of measurements are statistically independent and each segment measurement set S_i is strictly a function of the probability parameter p_i we can write:

$$f_{\mathbf{p}}\left(\mathbf{p}|\mathbf{n},\mathbf{N}\right)=\prod_{i=1}^{M} f_p\left(p_i|n_i,N_i\right)=\prod_{i=1}^{M}\beta^{-1}\left(n_i+1,N_i-n_i-1\right)p_i^{n_i}\left(1-p_i\right)^{N_i-n_i} \quad (5)$$

For clarity in the derivations below, the weighted estimate ℓ is notated as $\ell=\sum_i y_i$, where $y_i = w_i p_i$. Here y_i is a one-to-one transformation of p_i such that $p_i = y_i / w_i$. Since each p_i is Beta distributed, the distribution of y_i is also a Beta distribution:

$$f_y\left(y_i|n_i;N_i\right)=w_i^{-1}f_p\left(p_i/w_i|n_i;N_i\right), \quad (6)$$

where f_p (p_i / w_i | n_i ; N_i) is given by (3) upon substituting p_i / w_i for p_i. In addition, since the set of conditional estimates are statistically independent by virtue of (5), write:

$$f_\ell\left(\ell|n_1,N_1;\ldots;n_M N_M\right)=f_{y1}\left(\ell|n_1,N_1\right)\otimes\cdots\otimes f_{yM}\left(\ell|n_M,N_M\right), \quad (7)$$

where $\otimes$ denotes a convolution operation, i.e., $f\left(y\right)\otimes g\left(y\right)=\int f\left(y-x\right)g\left(x\right)dx$. Finally, the associated confidence C, analogous to (4), for the weighted estimate ℓ of regional system Pd can be expressed as:

$$C=\text{Prob}\left(\ell\geq p^*|n_1,N_1;\ldots,n_M,N_M\right)=\int_{p^*}^{1}f_\ell\left(\ell|n_1,N_1;\ldots,n_M,N_M\right)d\ell. \quad (8)$$

The probability density function f_ℓ (ℓ | n_1, N_1;…; n_M, N_M) is referred to as the *Augmented Beta Distribution*. Note that this is not a Beta distribution. The integral does not have a tractable closed form solution, but could be evaluated for specific parameter values through numerical methods.

The Augmented Beta Distribution f_ℓ (ℓ | n_1, N_1;…; n_M, N_M) is also appropriate when multiple target types are present within Test. This occurs when the weighted estimate ℓ for regional system Pd consists of a Beta distributed success probability p_i for each target type within each segment and f_ℓ (ℓ | n_1, N_1;…; n_M, N_M) results from a convolution across segment success probabilities p_i, which are themselves a convolution of target type success probabilities within the segment.

The Gaussian Approximation of System Pd

Returning to the weighted mean for system Pd (i.e., ℓ), assume that ℓ is approximately Gaussian distributed for a sufficiently large number of segments M. Because $\ell = \sum_i w_i p_i$, the corresponding mean $m_\ell(n_1, N_1; \ldots; n_M, N_M) = \sum_i w_i \mathrm{E}(p_i)$, where E(.) is the expectation operator. However, since p_i is Beta distributed, $\mathrm{E}(p_i) = (n_i + 1) / (N_i + 2)$ and the mean m_ℓ can be written as (Abramowitz & Stegun, 1965, p. 930):

$$m_\ell(n_1, N_1; \ldots; n_M, N_M) = \sum_{i=1}^{M} w_i (n_i + 1) / (N_i + 2). \tag{9}$$

Similarly, the associated variance is:

$$\sigma_\ell^2 = \sum_{i=1}^{M} w_i^2 \mathrm{Var}(p_i) \tag{10}$$

where $\mathrm{Var}(p_i) = (n_i + 1)(N_i - n_i + 1) / (N_i + 2)^2 (N_i + 3)$ (Abramowitz & Stegun, 1965, p. 930). Thus:

$$\sigma_\ell^2 = \mathrm{Var}(\ell | n_1, N_1; \ldots; n_M, N_M) = \sum_{i=1}^{M} \frac{(n_i + 1)(N_i - n_i + 1)}{(N_i + 2)^2 (N_i + 3)}. \tag{11}$$

Substituting these expressions for the mean and variance into the Standard Normal Distribution, allows us to compute an approximate $(1 - p^*)$ confidence interval for ℓ as:

$$\begin{aligned} C &= \left(2\pi\sigma_\ell^2\right)^{-1/2} \int_{p^*}^{1} \exp\left\{-(s - m_\ell)^2 / 2\sigma_\ell^2\right\} ds \\ &= F\left[(1 - m_\ell)/\sigma_\ell\right] - F\left[(p^* - m_\ell)/\sigma_\ell\right], \end{aligned} \tag{12}$$

where $F(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp\left(-s^2/2\right) ds$ is the Standard Normal Distribution.

The above assumption that the Augmented Beta Distribution converges to a Gaussian distribution for a large number of aggregate trials $\sum_i N_i$ holds when the following three conditions are satisfied:

I) The random variables y_i are independent and have finite mean μ_i and variance σ_i^2.

II) A raw moment greater than $2 + \delta$ is finite, i.e., $E(|y_i|^{2+\delta})$ is bounded for some $\delta > 0$ and for every $1 \le i \le M$.

III) $\lim_{M\to\infty} \frac{1}{S_M^{2+\delta}} \sum_{i=1}^{M} E\left(|y_i - \mu_i|^{2+\delta}\right) = 0$ for every $1 \le i \le M$ and for some $\delta > 0$ (known as Lyapunov's Condition), where $S_M^{2+\delta} = \left(\sum_{i=1}^{M} \sigma_i^2\right)^{(2+\delta)/2}$.

These three conditions describe the Lyapunov variant of the Central Limit Theorem. Proving that these conditions are satisfied for this problem will establish that the regional system sample mean is approximately Gaussian or normally-distributed. The details of this proof are given below. Recall by the previous definition that $y_i = w_i p_i$.

Proof:

I. It is easily verified that both μ_i and σ_i^2 are bounded from (9) and (10).

II.

a. $E(|y_i|^{2+\delta})$ being bounded implies that there exists a real number $R \le \infty$ such that $E(|y_i|^{2+\delta}) \le R$ for all $y_i = w_i p_i$

b. Letting $\delta = 2$, it can be shown that $E(|y_i|^4)$ is bounded, based on the use of the recursion relation $E\left(y_i^k\right) = \left(n_i + k\right) E\left(y_i^{k-1}\right) / \left(N_i + k + 1\right)$ and the fact that $E\left(y_i^2\right)$ is bounded (Johnson, Miller, & Freund, 1995, p. 586).

III.

a. It can be shown that $E\left(|y_i - \mu_i|^4\right) \le w_i^4$, since $0 \le E\left(|p_i - E(p_i)|^4\right) \le 1$, which follows from the fact that:

i. $y_i = w_i p_i$.

ii. $0 \le p_i \le 1$.

iii. $0 \le \mathrm{E}(p_i) \le 1$.

b. Additionally, because $S_M^2 = \sum_{i=1}^{M} \sigma_i^2$, as previously defined in condition III, conclude $S_M^2 \ge \sigma_0^2 \sum_{i=1}^{M} w_i^2$, when $\sigma_0^2 = \min_i \left[\mathrm{Var}\left(p_i \right) \right] > 0$, using the fact that there always exists a smallest *non-zero* Var (p_i) by virtue of (10).

c. Combining the results of the first two steps, i.e., $0 \le \mathrm{E}\left(\left| p_i - \mathrm{E}\left(p_i \right) \right|^4 \right) \le 1$ and $S_M^2 \ge \sigma_0^2 \sum_{i=1}^{M} w_i^2$, the expression $S_M^2 \sum_{i=1}^{M} \mathrm{E}\left(\left| p_i - \mathrm{E}\left(p_i \right) \right|^4 \right)$ is bounded by:

$$0 \le \frac{1}{S_M^4} \sum_{i=i}^{M} E\left(\left| y_i - \mu_i \right|^4 \right) \le \frac{1}{\sigma_0^4} \frac{\sum_{i=1}^{M} w_i^4}{\left(\sum_{i=1}^{M} w_i^2 \right)^2}. \qquad (13)$$

This bound implies that if $\lim_{M \to \infty} \frac{\sum_{i=1}^{M} w_i^4}{\left(\sum_{i=1}^{M} w_i^2 \right)^2} = 0$, then the following limit $\lim_{M \to \infty} \frac{1}{S_M^4} \sum_{i=1}^{M} E\left(\left| y_i - \mu_i \right|^4 \right) = 0$, which allows us to conclude that condition III is satisfied.

d. Without loss of generality, assume that the weights are bounded above and below by w_U and w_L, respectively, such that $0 \le w_L \le w_i \le w_U \le 1$ Even if there did exist a finite number of zero weights, the remaining weights could be re-indexed so that $w_L > 0$. Now, it follows that $\sum_{i=1}^{M} w_i^4 \le M w_U^4$ and $\sum_{i=1}^{M} w_i^2 \ge M w_L^2$. This implies that $\sum_{i=1}^{M} w_i^4 \Big/ \left(\sum_{i=1}^{M} w_i^2 \right)^4 \le \left(w_U / w_L \right)^4 \Big/ M$. From (13), conclude that the Lyapunov condition is satisfied since

$\lim_{M\to\infty} \frac{1}{S_M^4} \sum_{i=1}^{M} E\left(\left|y_i - \mu_i\right|^4\right) = 0$ and condition III is satisfied when there are an infinite number of *nonzero but bounded* weights.

e. For the case where all the weights are uniform, $w_i = 1 / M$; $1 \le i \le M$, it follows from the results in (d) that condition III is again satisfied.

By the Lyapunov variant of the Central Limit Theorem, conditions I–III being true imply that the sum or average of y_i is Gaussian for large M and that the regional system sample means for Pd are Gaussian. However, for the highly unlikely cases where all but a small number of arrival weights w_i are zero, then the right-hand side of (13) may no longer be zero in the limit of large M resulting in the Gaussian approximation becoming invalid. Intuitively though, it can be reasoned that this case is not possible, because a segment of an operationally valid route cannot have a zero probability of being traversed by an item-of-interest. This would be especially true for a segment selected for test.

Example of the Augmented Beta and its Gaussian Approximation

As an illustration, consider the following hypothetical example where a weighted detection probability estimate is constructed from ten independent segments. The number of trials per segment for this example is 5, 10, 8, 6, 9, 7, 4, 5, 8, and 9, respectively. Note, however, that the number of trials per segment has no impact on the convergence to Normality. The corresponding number of successful detections declared is 4, 8, 7, 5, 7, 5, 2, 4, 6, and 8, respectively. In addition, the probability associated with choosing a given segment is 0.2, 0.1, 0.3, 0.05, 0.1, 0.025, 0.15, 0.025, 0.025, and 0.025, respectively. Using (7), the Augmented Beta Distribution can be computed and is of the form illustrated in Figure 3. The convolutions are approximated discretely for a step size $r = 0.0001$ so that the integrated density in the interval [0, 1] is nearly unity. The distribution is shown to be both uni-modal and approximately symmetric about its mean value (~ 0.7). In addition, the distribution can be well-approximated by a Gaussian distribution–it passes both the Kolmogorov-Smirnov and Chi-Square Goodness-of-Fit tests.

This illustration suggests that the Augmented Beta Distribution converges to the Gaussian distribution using a batch of 10 segments. Therefore, it is reasonable to infer that the regional system sample mean, which is going to be obtained from a batch of N segments with $N > 30$, will rapidly converge to the Gaussian distribution. The proof provided above that as M gets large, the regional system sample mean becomes Gaussian is borne out in this illustration which suggests that this convergence begins to occur when $M = 10$.

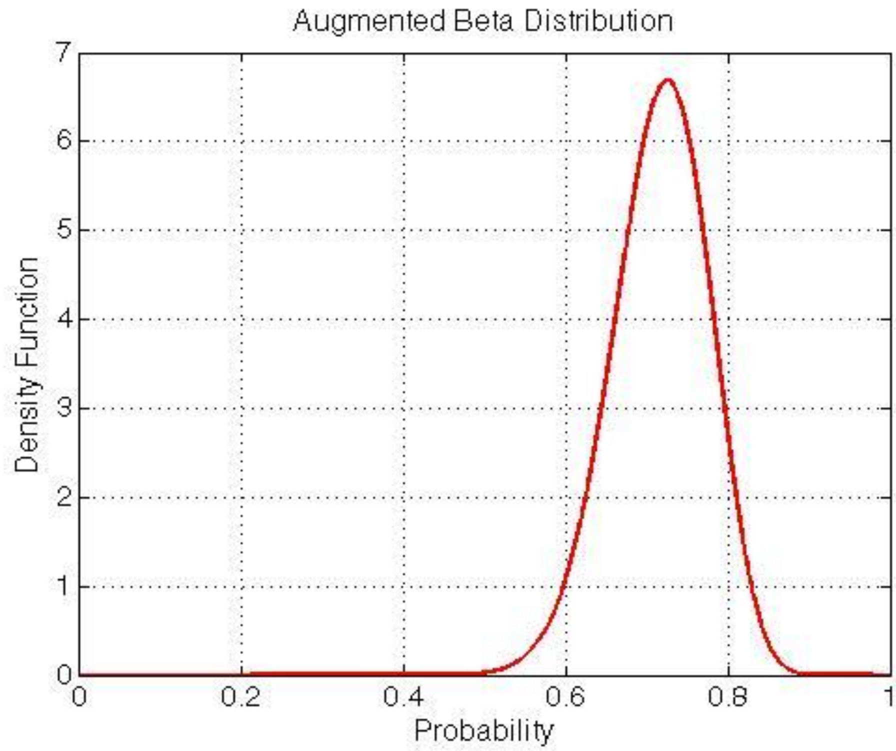


Figure 3. The Augmented Beta Distribution corresponding to 10 independent segments, where the number of hypothetical trials per segment is N = [5, 10, 8, 6, 9, 7, 4, 5, 8, 9], the number of detections is n = [4, 8, 7, 5, 7, 5, 2, 4, 6, 8] and the associated probability of choosing each segment is q = [0.2, 0.1, 0.3, 0.05, 0.1, 0.025, 0.15, 0.025, 0.025, 0.025].

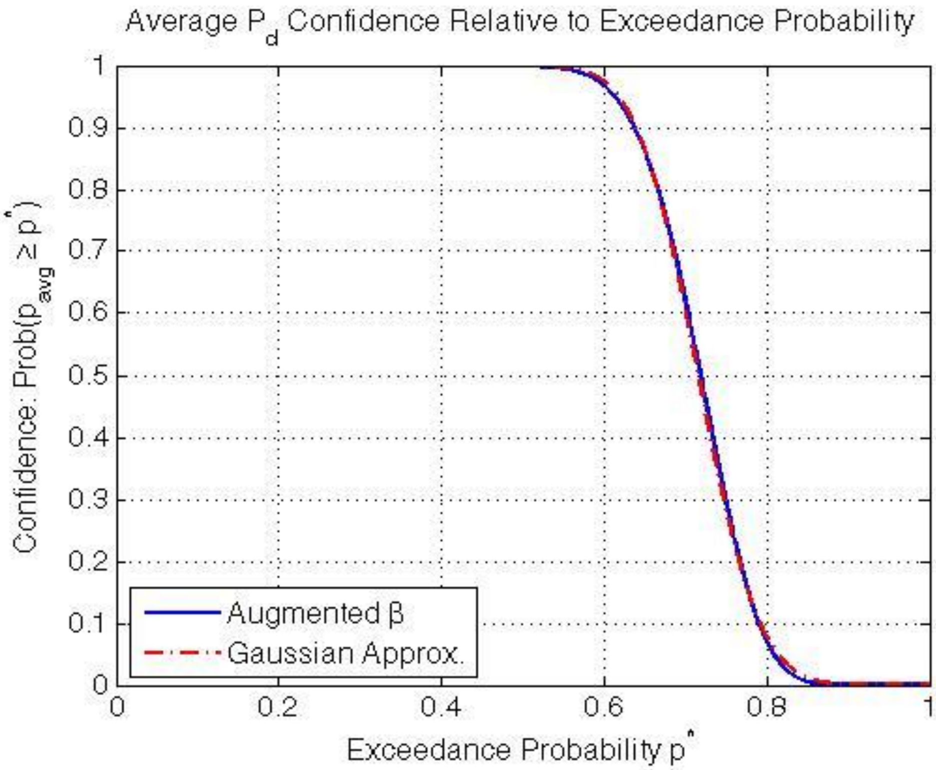


Figure 4. The confidence *C* for the Augmented Beta Distribution associated with the weighted detection probabilities for 10 independent segments, where the number of hypothetical trials per segment is *N* = [5, 10, 8, 6, 9, 7, 4, 5, 8, 9], the number of detections is *n* = [4, 8, 7, 5, 7, 5, 2, 4, 6, 8] and the associated probability of choosing each trial is *q* = [0.2, 0.1, 0.3, 0.05, 0.1, 0.025, 0.15, 0.025, 0.025, 0.025].

Figure 4 is a plot of the weighted detection probability confidence C as a function of the exceedance probability p^* for both the Augmented Beta Distribution and its corresponding Gaussian approximation. It is apparent that both distribution functions are nearly identical. This is true when the aggregate number of trials $\sum_i N_i$ is sufficiently large. For a confidence $C = 0.90$, the exceedance probability is approximately 0.65 for either the exact or Gaussian approximation.

Extension to Pid

The previous analysis for Pd presented above assumed that each segment i has a single, but unknown success probability p_i representative of the entire segment.

Consider the case where a specific segment i consists of multiple distinct and unknown success probabilities $p_{i,j}$, where $j = 1, 2,\ldots, J$. This situation is relevant to the identification problem for a given segment, where a single success probability p is not representative of the probability of successful or correct identification across a whole segment because multiple distinct success probabilities $p_{i,j}$ exist. Our reasoning is based on the argument that the window of processing for an identification trial is much smaller than for a detection trial. In particular, the window for detection extends across the entire segment. Because identification is a human-driven action comprised of multiple concurrent processing tasks, the trial window cannot possibility extend across the whole segment. Therefore, identification attempts within a segment can occur at different locations where the system attributes can vary. This implies that the success probabilities between any two identification trials, even within a segment, cannot be assumed to be equal.

Consider how the analysis above can be extended to address Pid, where multiple distinct and unknown success probabilities $p_{i,j}$ occur within a segment. Similar to the derivation above, assume that each segment consists of N_i trials but now n_i corresponds to successfully declared identifications. The number of successful identifications can be characterized by $n_i = \sum_{n=1}^{N_i} \chi_{n,i}$, where $\chi_{n,i} = \{0,1\}$ is an indicator function that represents an incorrect or correct identification, respectively. It is assumed that $\chi_{n,i}$ can be drawn from a mixture distribution of the form:

$$\chi_{n,i} \sim \begin{cases} B\left(x,1;p_{1,i}\right) & \text{with probability } q_{1,i} \\ B\left(x,1;p_{2,i}\right) & \text{with probability } q_{2,i} \\ & \vdots \\ B\left(x,1;p_{j,i}\right) & \text{with probability } q_{\mathrm{j},i} \end{cases}, \tag{14}$$

where $B\,(x, 1, p_{j,i})$ is a binomial distribution for a single trial subinterval within a segment i, having success probability $p_{j,i}$, $x = \{0, 1\}$ and $q_{j,i}$ is the probability that $\chi_{n,i}$ is drawn from the distribution characterizing subinterval j on segment i. Let p_i be the probability associated with the random variable $\chi_{n,i}$ such that:

$$p_i = \text{Prob}\left(\chi_{n,i}\right) = \sum_{j=1} q_{j,i} p_{j,i} \tag{15}$$

Given that each trial subinterval is independent of any other trial (naturally resulting since the location window varies for each identification trial) and the mixture is uniform across all subintervals, n_i is binomially-distributed with average success probability p_i. For this analysis, it is not necessary to know the probabilities $p_{j,\,i}$ and $q_{j,\,i}$. From this result, it now follows that the Augmented Beta Distribution can be used for Pid and the Lyapunov convergence proof outlined above is valid for this generalization provided that the exceedance probability is interpreted as the *average system identification performance* along a given segment. The exact expression for the confidence C in (8) and its Gaussian approximation (*cf.* (12)) can then be used without modification. Moreover, the mixture distribution characterization also allows for a relaxing of the assumption that the success probabilities for detection must be equal for trials within a segment. This suggests that these confidence intervals can also be used on test designs consisting of multiple target types.

The following example illustrates how a mixture distribution of uniform subintervals within a segment with varying success probabilities results in the average number of successes on the segment being binomially distributed. Suppose there are 10 independent trials ($N_i = 10$) along a given segment i, where the unknown success probabilities are $p_{j,\,i} = [0.9, 0.7, 0.4, 0.8]$ and the occurrence probabilities associated with these success probabilities are $q_{j,\,i} = 0.25$ for $j = 1, 2,\ldots,4$. Figure 5 depicts the distribution function for the number of successful detection attempts resulting from a Monte-Carlo sampling of the mixture distribution as defined previously in (15) with the $p_{j,\,i}$ and $q_{j,\,i}$ values noted above.

The blue bars represent the histogram resulting from the Monte-Carlo sampling with the theoretical binomial distribution (red curve) overlaid using the average success probability defined in (15). The results show excellent agreement between the two distributions and justify the use of applying the Gaussian approximation results for Pd to Pid.

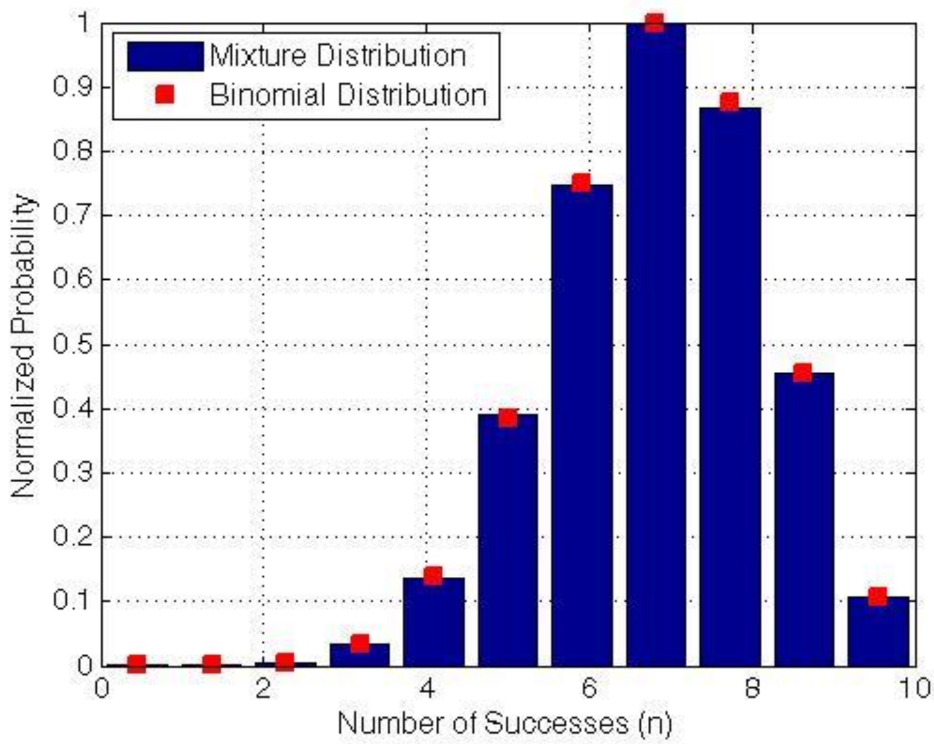


Figure 5. A comparison of the Monte Carlo-based distribution function and the theoretical binomial distribution for the following mixture distribution: $p_{j,i}$ = [0.9, 0.7, 0.4, 0.8] and $q_{j,i}$ = 0.25 for j = 1, 2,…,4, where N_i = 10.

Confidence Interval Equations and Sample Calculations

Consider the equations for constructing the confidence interval for the system mean Pd and Pid, as verified in the analysis provided above. The method generates the probability or confidence that the regional system sample mean exceeds a given threshold and is obtained under a Normal approximation when the segment success probabilities p_i are Beta distributed. The equations for constructing the confidence interval are given below and are followed by a numerical example depicting their implementation on hypothetical data.

Given that the distribution of regional system sample mean is approximately Normally distributed and the segment success probabilities p_i are Beta distributed, the following equation is used to compute the probability that the system Pd and Pid mean exceeds a given threshold p^* with some confidence C:

$$
\begin{aligned}
C &= \left(2\pi\sigma_\ell^2\right)^{-1/2} \int_{p^*}^{1} \exp\left\{-\left(s-m_\ell\right)^2 / 2\sigma_\ell^2\right\} ds \\
&= F\left[\left(1-m_\ell\right)/\sigma_\ell\right] - F\left[\left(p^* - m_\ell\right)/\sigma_\ell\right],
\end{aligned}
$$

where

$$
\begin{aligned}
&F(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp\left(-s^2/2\right) ds, \\
&m_\ell = \sum_{i=1}^{M} w_i \left(n_i+1\right)/\left(N_i+2\right), \quad \sigma_\ell^2 = \sum_{i=1}^{M} w_i^2 \frac{\left(n_i+1\right)\left(N_i-n_i+1\right)}{\left(N_i+2\right)^2\left(N_i+3\right)}
\end{aligned}
$$

N_i and n_i are the total number of detection attempts and successfully declared detections in SAT, respectively, for each segment i.

The following example shows how the above equation for Pd and Pid confidence intervals can be implemented using hypothetical test results consisting of observed detections and identifications distributed across non-identical segments within a region. The hypothetical test involves 56 segments with 5 potential trials occurring on each segment for detection and identification. The reader should note here that each segment may involve any mixture of target types since the proposed methodology is valid under any target type configuration supported by the program's current experimental design. Table 1 summarizes the hypothetical detection and identification observations.

A '1' appearing in Table 1 denotes a successful detection or identification while a '0' represents an unsuccessful attempt. An 'x' labeled within the identification columns indicates that a trial is not counted due to an unsuccessful detection.

The number of detection trials observed during this hypothetical test is 280 with 224 detections successfully declared. This, therefore, results in 224 potential identifications. The sample mean for Pd is simply 224 / 280 = 0.80 and similarly the sample mean for Pid is 202 / 224 = 0.90.

Table 1. Hypothetical Test Results

Segment	Detection Trial					Identification Trial				
	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
1	1	1	1	1	1	1	1	0	1	1
2	1	1	0	1	1	1	1	x	1	1
3	1	1	0	1	1	1	1	x	1	1
4	1	1	1	1	0	1	1	1	1	x
5	1	1	1	1	1	1	1	0	1	1
6	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	x	1	1	1	1
8	0	1	1	1	1	0	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	0	1	0	1	1	x
13	0	0	1	1	1	x	x	1	1	1
14	1	1	1	0	0	1	1	1	x	x
15	1	1	1	0	0	1	1	1	x	x
16	1	1	1	1	1	1	1	1	1	1
17	1	1	1	1	1	1	1	0	1	1
18	1	1	1	1	1	0	1	1	0	1
19	1	0	0	1	1	0	x	x	1	1
20	1	1	1	1	1	0	1	1	1	1
21	1	1	1	1	1	1	1	1	1	1
22	0	1	1	1	1	x	1	1	1	1
23	1	1	1	1	1	1	1	1	1	1
24	1	0	1	0	1	1	x	1	x	1
25	1	1	1	1	1	1	1	1	1	1
26	1	1	0	1	1	1	1	x	1	1
27	0	1	0	1	1	x	1	x	1	1
28	0	1	0	1	1	x	1	x	1	1
29	1	0	1	1	1	1	x	1	1	1
30	1	1	1	1	1	1	0	1	1	1
31	1	1	1	1	0	1	1	0	1	x
32	1	1	0	1	1	1	1	x	1	1
33	1	1	0	1	1	1	1	x	1	1
34	0	1	1	1	0	x	1	1	1	x
35	1	1	1	1	1	1	1	0	1	1
36	1	1	1	1	1	1	1	1	1	0
37	1	1	1	0	0	x	1	1	x	x
38	0	1	1	1	1	0	1	1	1	1
39	1	0	0	1	1	1	x	x	1	1
40	1	1	1	1	1	1	1	1	1	1
41	1	1	1	1	1	1	1	1	1	1
42	1	1	0	1	1	1	0	x	1	1
43	0	0	1	1	1	x	x	1	1	1
44	1	0	1	1	1	1	x	1	1	1
45	1	0	0	1	1	1	x	x	0	1
46	0	0	0	0	0	x	x	x	x	x
47	1	1	1	1	0	1	1	0	1	x
48	1	0	1	1	1	0	x	1	0	0
49	1	0	0	1	1	0	x	x	1	1
50	0	1	1	0	0	x	1	1	x	x
51	1	1	1	1	1	1	1	1	1	1
52	0	1	1	1	0	x	1	1	1	x
53	1	1	1	1	1	1	1	1	1	1
54	1	0	1	0	1	1	x	1	x	1
55	1	1	1	1	1	1	1	1	1	0
56	0	1	0	1	1	x	1	x	1	1

To compute the confidence on the Pd and Pid mean, first compute the first and second central moments from (9) and (11), respectively, for both Pd and Pid:

$$m_l = \sum_{i=1}^{M} w_i \left(n_i + 1\right) / \left(N_i + 2\right), \quad \sigma_l^2 = \sum_{i=1}^{M} w_i^2 \frac{\left(n_i + 1\right)\left(N_i - n_i + 1\right)}{\left(N_i + 2\right)^2 \left(N_i + 3\right)}$$

These calculations for the first and second central moments are shown separately in the following two sub-sections for the Pd and Pid confidence intervals. The remaining steps required to establish that the regional sample mean Pd or Pid is greater than p^* are also provided within each subsection.

Confidence of Pd > p^*

Because w_i = 1/30 due to a necessary assumption of uniform threat arrival weights, the following calculations can be performed for m_l and σ_l^2, respectively:

$$m_l = \sum_{i=1}^{M} w_i \left(n_i + 1\right) / \left(N_i + 2\right) = \sum_{i=1}^{30} \left(n_i + 1\right) / 210 = 0.7145$$

$$\sigma_l^2 = \sum_{i=1}^{M} w_i^2 \frac{\left(n_i + 1\right)\left(N_i - n_i + 1\right)}{\left(N_i + 2\right)^2 \left(N_i + 3\right)} = 0.0004$$

Converting the above values for mean and variance into a standardized random variable z for an assumed exceedance probability $p^* = 0.7$:

$$z_1 = \left(p^* - m_l\right) / \sigma_l = \left(0.7 - 0.7145\right) / \sqrt{0.0004} = -0.7245,$$

and

$$z_2 = \left(1 - m_l\right) / \sigma_l = \left(1 - 0.7145\right) / \sqrt{0.0004} = 14.2751,$$

with z having the Standard Normal Distribution: $F\left(x\right) = \left(2\pi\right)^{-1/2} \int_{-\infty}^{x} \exp\left(-s/2\right) ds$.

Then, through the use of a standard look-up table for the Standard Normal Distribution (Johnson, Miller, & Freund, 1994, p. 586), the probability that the mean system Pd is greater than p^* is F (14.2751) – F (-0.7245) = 0.7656.

*Confidence of Pid > p**

Again, since $w_i = 1/30$, the following calculations can be performed for m_l and σ_l^2, respectively:

$$m_l = \sum_{i=1}^{M} w_i \left(n_i + 1\right) / \left(N_i + 2\right) = \sum_{i=1}^{30} \left(n_i + 1\right) / \left(30\left(N_i + 2\right)\right) = 0.8044$$

$$\sigma_l^2 = \sum_{i=1}^{M} w_i^2 \frac{\left(n_i + 1\right)\left(N_i - n_i + 1\right)}{\left(N_i + 2\right)^2 \left(N_i + 3\right)} = 0.0009.$$

Note that for Pid, the values of N_i change for each i since the number of identification trials per segment is dependent on the number of successful detections for that segment. Similarly, converting the above values for mean and variance into a standardized random variable z for an assumed exceedance probability $p^* = 0.7$:

$$z_1 = \left(p^* - m_l\right) / \sigma_l = \left(0.7 - 0.8044\right) / \sqrt{0.0009} = -3.480,$$

and

$$z_2 = \left(1 - m_l\right) / \sigma_l = \left(1 - 0.8044\right) / \sqrt{0.0009} = 6.5200.$$

Then, through the use of a standard look-up table for the Standard Normal Distribution, the probability that the regional system mean is greater than p^* is F (6.5200) – F (-3.480) = 0.9997. In summary, these example calculations show that there is a 76.56% and a 99.97% statistical confidence that the true system Pd and Pid mean is above 0.7, respectively, for this hypothetical set of test observations. It is apparent from the above example that more than five trials would be beneficial if the sample mean is within 0.1 of p^*. The difference in the hypothetical test sample means of 0.80 for Pd and 0.90 for Pid versus their Bayesian posterior expected values of 0.7145 and 0.8044, respectively, illustrate the need for additional trials for each segment. Moreover, this example showed that sample means around 0.80 only resulted in less than 80% confidence that the true mean is above 0.7.

Monte Carlo Confidence Interval Projector & Implementation

The evaluation of the various candidate test designs for this regional test involved evaluating projected confidence interval widths and exceedance probabilities. Moreover, there is a motivation to more exactly understand the relationship between the number of segments and test trials and confidence bound widths. This understanding would facilitate the decision of selecting a design with the fewest number of trials while still maintaining a strong likelihood in achieving a specified and desired confidence width. To expedite this analysis, a Monte Carlo confidence interval tool (CI Projector) is developed to automate the calculation of the confidence intervals derived above.

The CI Projector tool is coded within an Excel environment using VBA. It requires the user to input a candidate test design through specifying the number of routes, the number of segments, and the number of trials per segment by filling out the blue columns titled ‘Route’, ‘# of segments’, and ‘Ni’ as shown in Figure 6. Note that Ni simply refers to the number of segments on a route.

Figure 6 also illustrates the view from the model during execution. The five columns in the middle denoted ‘Segment successes’ are the sampled number of successes for each segment during each iteration. Sample mean and variance values for the segments, routes, and area are tallied and averaged after each iteration. This also provides the user with a subjective understanding of the amount of variability present in the confidence interval widths from run to run.

PD										
Route		Segment successes								
	# of segments	#1	#2	#3	#4	#5	Ni	Sr	Nr	Route avg
1	1	4					7	4	7	0.571428571
2	1	3					7	3	7	0.428571429
3	1	3					7	3	7	0.428571429
4	1	1					7	1	7	0.142857143
5	1	6					7	6	7	0.857142857
6	2	5	4				7	9	14	0.642857143
7	1	6					7	6	7	0.857142857
8	1	6					7	6	7	0.857142857
9	1	5					7	5	7	0.714285714
10	3	5	5	4			7	14	21	0.666666667
11	2	3	2				7	5	14	0.357142857
12	2	6	5				7	11	14	0.785714286
13	4	2	2	3	5		7	12	28	0.428571429
14	2	4	2	3			7	9	14	0.642857143
15	3	6	6	2			7	14	21	0.666666667
16	2	5	5				7	10	14	0.714285714
17	2	3	5				7	8	14	0.571428571
18	1	3					7	3	7	0.428571429
19	2	3	5				7	8	14	0.571428571
20	2	5	4				7	9	14	0.642857143
21	2	4	5				7	9	14	0.642857143
22	1	4					7	4	7	0.571428571
23	2	5	5				7	10	14	0.714285714
24	1	7					7	7	7	1
25	2	7	7				7	14	14	1
26	2	6	7				7	13	14	0.928571429
27	2	5	7				7	12	14	0.857142857
28	3	4	6	4			7	14	21	0.666666667
29	1	7					7	7	7	1
30	2	4	2				7	6	14	0.428571429
31	2	4	4				7	8	14	0.571428571
32	2	6	7				7	13	14	0.928571429
33	3	7	6	6			7	19	21	0.904761905
34	2	4	5				7	9	14	0.642857143

Figure 6. CI Projector Screenshot

The success distribution for each segment is of course dependent on the projected sample mean for Pd and Pid. The user also inputs the projected sample means for Pd and Pid in addition to the number of Monte Carlo replications to be performed as shown below in Figure 7. Sample means for Pd must be defined for each object type passing through the system, and for adverse weather. The projected Pid sample mean strictly represents an average over all of these components.

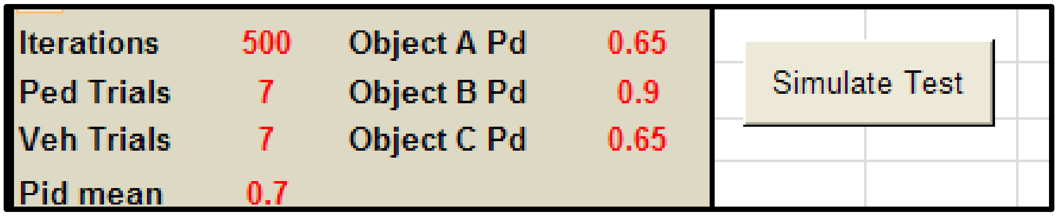
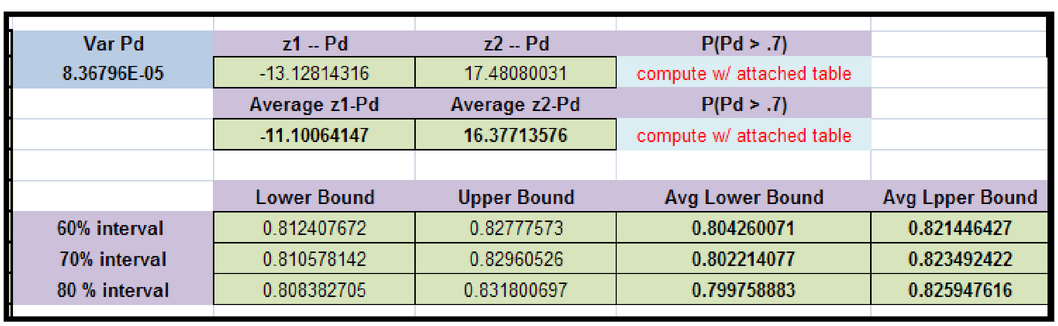


Figure 7. CI Projector Inputs

After defining the specific test design within the tool, the number of iterations must also be assigned by the user. The default value is 500, but the user can adjust this value to any number. Due to the fast speed at which CI Projector runs coupled with the existence of a small amount of variability within the resulting confidence interval widths , a high number of iterations is recommended and should always be used. Lastly, the user simply clicks on the 'Simulate Test' button as shown to execute the model.

Sample outputs for the test projector tool are provided in Figure 8. Provided in the upper portion of the output columns are the necessary numerical values to compute the exceedance probability for a threshold value of interest that can be easily inferred through an attached table. The user also directly obtains lower and upper bounds for 60%, 70%, and 80% confidence intervals.



Var Pd	z1 -- Pd	z2 -- Pd	P(Pd > .7)	
8.36796E-05	-13.12814316	17.48080031	compute w/ attached table	
	Average z1-Pd	Average z2-Pd	P(Pd > .7)	
	-11.10064147	16.37713576	compute w/ attached table	
	Lower Bound	Upper Bound	Avg Lower Bound	Avg Lpper Bound
60% interval	0.812407672	0.82777573	0.804260071	0.821446427
70% interval	0.810578142	0.82960526	0.802214077	0.823492422
80 % interval	0.808382705	0.831800697	0.799758883	0.825947616

Figure 8. Sample Output Confidence Intervals from CI Projector

This process of inputting candidate test designs, running the model, and observing resulting confidence interval widths and exceedance probabilities can be easily repeated to evaluate a wide range of test designs. The CI Projector tool provides an environment conducive to short scenario set-up time and run time.

Numerous designs are evaluated using this tool, giving the decision maker reasonable assurance that the desired confidence interval widths and exceedance probabilities will be achieved as a result of an efficient test design.

Conclusion

The probability density function (pdf) for regional system sample means is derived by considering them as a weighted estimate of the successful detection probability $\sum_i w_i p_i$ under a number of independent but *not* necessarily identically-distributed Bernoulli trials. It was shown that the resulting distribution (Augmented Beta Distribution) for $\sum_i w_i p_i$ is a convolution of non-identical Beta Distributions (*cf.* (7)). From this result, the corresponding confidence that the weighted estimate exceeds a given exceedance probability can be determined exactly from the Augmented Beta Distribution expression (*cf.* (8)). Then, the results were extended to address the regional system Pid sample mean. The analysis for the Pid sample means was more complicated and was based on the use of mixture distributions for cases where the success probability is no longer constant across a segment. It was shown that the same Pd results apply here provided that the exceedance probability is interpreted as the average system exceedance probability across the entire segment.

It was also shown through the Lyapunov variant of the Central Limit Theorem that the Augmented Beta Distribution converges to a Gaussian distribution as the number of segments grows large. The mathematical proof supplied showed that the Lyapunov conditions are satisfied for uniform segment priors w_i. Given this result, a simpler Gaussian approximation (*cf.* (12)) can be used to compute the confidence for both the regional system Pd and Pid sample mean that involves a simple aggregation of the detection and identification events observed across all segments during Test. These confidence intervals can be computed for individual target types and for a regional system average of all target types.

References

Abramowitz, M. & Stegun, I. A. (1965a). *Handbook of mathematical functions: With formulas, graphs, and mathematical tables*. New York, NY: Dover.

Bernardo, J. M. & Smith, A. F. (2000). *Bayesian theory*. New York, NY: Wiley. p. 241-255.

Box, G. E. P., Hunter, J. S. & Hunter, W. G. (2005). *Statistics for experimenters* (2nd ed.). New York, NY: John Wiley and Sons.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995). *Continuous univariate distributions* (Vol. 2). New York, NY: Wiley.

Johnson, R. A., Miller, I., & Freund, J. E. (1994). *Miller and Fruend's probability and statistics for engineers*. Englewood Cliffs, NJ: Prentice Hall. p. 586.

Karr, A. F. (1993). *Probability*. New York, NY: Springer-Verlag. p. 190-192.