STAT COE-Report-08-2017

Binary Response and Single Stress Factor Test Methods

Authored by: Sarah Burke, PhD
Lenny Truett, PhD

15 June 2017

Revised 29 August 2018

The goal of the STAT COE is to assist in developing rigorous, defensible test strategies to more effectively quantify and characterize system performance and provide information that reduces risk. This and other COE products are available at www.afit.edu/STAT.

STAT Center of Excellence
2950 Hobson Way – Wright-Patterson AFB, OH 45433

Table of Contents

Revision 1, 29 Aug 2018: formatting and minor typographical/grammatical edits

Executive Summary

The Department of Defense (DoD) is often interested in the ballistic resistance properties of various types of armor. The goal of these experiments may be to characterize the ballistic response curve or to estimate a specific velocity on the response curve. We describe four test methods for this type of sensitivity experiment and discuss their advantages and disadvantages. We highlight a more recent sequential method, the three phase optimal design, which has been shown to provide efficient estimates of the ballistic response curve in an affordable number of runs.

Keywords: ballistic resistance, sensitivity testing, three phase optimal design, sequential experimentation, logistic regression

Introduction

The Department of Defense (DoD) is often interested in the ballistic resistance properties of various types of armor. In this type of experiment, a bullet or projectile is shot at the armor with a given velocity. The shot results in either a complete penetration (CP) or partial penetration (PP) where non-penetrations are classified as PPs. The goal of these experiments may be to characterize the ballistic response curve or to estimate a specific velocity on the response curve such as V_{50} or V_{10}, the velocity at which there is a 50% and 10% probability of CP, respectively. This is an example of a sensitivity experiment where the goal is to understand the effect of a single stress factor (e.g., velocity) on a binary outcome (e.g., penetration of armor).

This best practice examines four methodologies for these types of sensitivity experiments: lot acceptance sampling plans, the up-down method, the Langlie method and the three phase optimal design (tomsshinyapps). Although we highlight these methods as applied to a ballistic resistance test, they have many applications across various fields of study. For example, in clinical trials a sensitivity experiment is used to determine the maximum tolerated dose of a drug. The experimenters assume that benefits to the patient may increase as the dose of the drug increases, but the toxicity increases as well. An experiment is performed on a small set of patients where the dose is methodically changed to determine the maximum tolerated dose. Another medical application is determining the probability of detection for a substance in a patient's blood as a function of concentration. Reliability engineers use this same type of experiment to determine the probability of product or system failure as a function of the number of cycles, pressure, or yield load.

Ballistic resistance testing is a destructive, often expensive test. A test strategy that can successfully estimate the ballistic response curve and/or a particular point along this curve in an affordable number of runs is essential. In the next several sections, we discuss the advantages and disadvantages of four test strategies for sensitivity experiments. We conclude with a summary of many advantages of a more recent method, three-phase optimal design (tomsshinyapps) developed by Wu and Tian (2013).

Lot Acceptance Sampling Plans

Lot acceptance sampling plans (LASPs) are typically used to determine acceptance or rejection of a lot. Because it is inefficient and/or costly to test every item in the lot, a random sample of items from the lot are inspected or tested to determine whether to accept or reject the entire lot. Acceptance sampling plans are defined by the sample size, n, and the acceptance number (i.e., number of allowed failures), c. The user must specify an acceptable quality limit, rejectable quality limit, consumer's risk, and producer's risk to determine the sampling plan. Refer to Truett (2013) and Harman (2013) for a more thorough discussion of acceptance sampling plans.

In the context of ballistic resistance testing, an acceptance sampling plan may be used to verify V_{50}, V_{10}, or another quantile of interest V_p. Suppose you believe the velocity with 10% probability of CP is some value V. A sampling plan with n runs and acceptance number c is used to test the probability of CP at the tested velocity. If there are c or fewer CPs in n trials, the initial assumption (e.g., that V_{10} = V) is accepted; if there are more than c CPs, the assumption is rejected. Only one point on the response curve is tested using this approach (e.g. V_{10}); however, an acceptance sampling is not done to *estimate* a specific point on the curve. This method simply assesses performance at a particular velocity.

The assumptions in the analysis for this method are that the response is binary (i.e., pass/fail), the test size is fixed at n, each run in the test is independent, and the probability of penetration is constant for each run in the test. The first three assumptions are typically met in ballistic resistance tests; however, the last assumption is likely violated because there are multiple conditions necessary for it to hold. One of the critical assumptions in LASP is that each item in the lot has been produced exactly like the others under highly controlled conditions. In ballistic resistance testing, this assumption may not hold because of the inherent variability in velocity of the projectile.

LASPs have a few advantages: they are very simple to implement and the results lead to a conclusive decision of pass or fail. However, acceptance sampling plans have several drawbacks in addition to the potential assumption violation previously discussed. Suppose we are interested in assessing V_{10}. A test with 80% confidence yields a LASP with sample size $n = 16$ and acceptance number $c = 0$. Figure 1 shows an operating characteristic curve for this plan. If we observe one (or more) CPs, the test fails and we have gained limited information. We do not know at which velocity the true V_{10} is. Now suppose we perform the test and do not observe a CP, that is we have "passed" the test. We still cannot conclude that the tested velocity is the true V_{10}; the true V_{10} may in fact be a value larger than the velocity tested. In addition to the limited conclusions we can draw from this type of test, these methods also require large sample sizes to have sufficient confidence and power. With limited test resources in ballistic resistance testing, it is often inefficient and costly to use acceptance sampling plans. However, LASPs may be appropriate when determining whether to accept a lot of armor. In these cases, it is likely not necessary to test the armor at several different values.

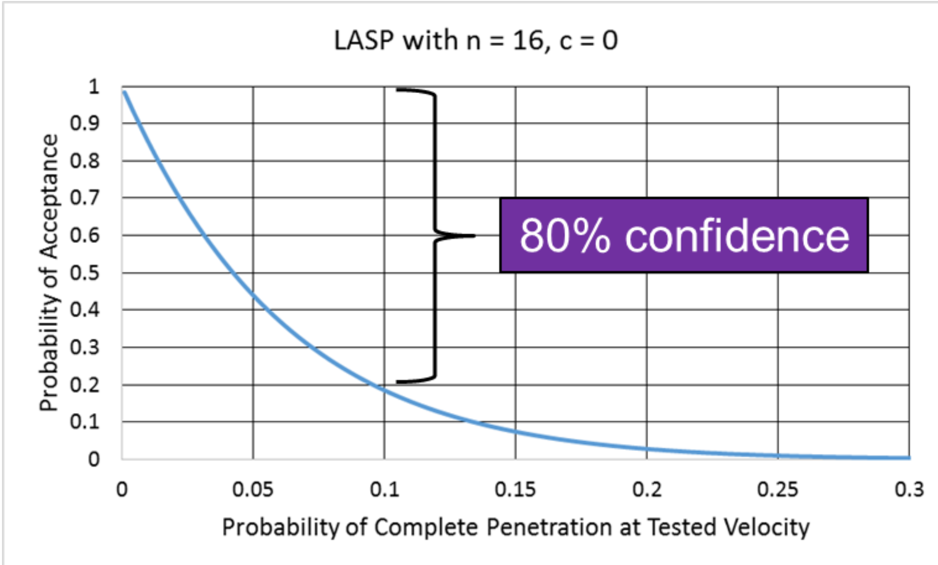

Figure 1. Operating characteristic curve for acceptance sampling plan with n = 16 and c = 0

In the following sections, we discuss three sequential test methods: the up-down method, the Langlie method, and 3pod. The key assumptions utilized for these methods are that there is a binary response and a single stress factor. As the stress factor value increases, the probability of the response occurring increases. A response curve can be estimated for the probability of CP using a logistic regression model. Similar to a linear regression model, a logistic regression model provides an estimate for the probability of a success (e.g., CP) given values of a factor (e.g., velocity). For one factor x and a binary response y, the model takes the form:

$$P(y = CP|x) = \frac{1}{1 + e^{-(\beta_0+\beta_1 x)}} \tag{1}$$

The model parameters β_0 and β_1 can be readily estimated by statistical software using maximum likelihood. Using this model, confidence intervals (CIs) can also be calculated for a quantile V_p on the ballistic response curve (e.g., V_{10} or V_{50}). For more information on logistic regression, refer to Myers et al. (2010).

One concern when modeling a binary response as a function of one or more input variables is the issue of separation. Figure 2 shows an example of separation for one input factor. In Figure 2a , there is no overlap in the responses so that when $X < 9$, the response is always a fail and for $X > 9$, the response is always a pass. When there is no overlap, separation occurs, and the response curve cannot be estimated, resulting in limited analysis options. In Figure 2b, there is overlap in responses (i.e., no separation) as indicated by the vertical line (not all passes are on one side of the line). When there is no separation as in Figure 2b, the response curve, and consequently any quantile, is estimable. The following methods for sensitivity experiments handle this issue of separation differently.

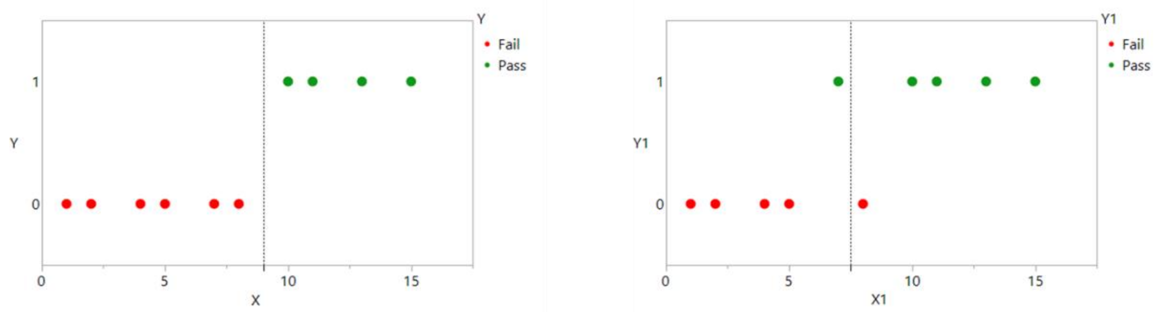

Figure 2. Illustration of Separation: In a) the response curve cannot be estimated. In b) the overlap in responses allows response curve to be estimated.

Up-Down Method

One common method to estimate V_{50} is the up-down method as described in MIL-STD-662F (1997). This method requires an initial guess on the velocity μ_G where the estimated probability of penetration is 50%. The test points are chosen as follows:

1. $x_1 = \mu_G$
2. If y_1 is a CP, $x_2 = x_1 - \delta$ ft/s
   If y_1 is a PP, $x_2 = x_1 + \delta$ ft/s.
3. The velocity is increased or decreased δ ft/s until one PP and CP are observed.
4. Once a PP and CP are observed, $x_{i+1} = x_i + \delta$ ft/s if y_i is a PP.
   If y_i is a CP, $x_{i+1} = x_i - \delta$ ft/s.

The stopping criterion for this method occurs when an equal (pre-specified) number k of PPs and CPs have been observed (typically 3 or 5). V_{50} is then estimated as the arithmetic average of the k highest velocities resulting in PPs and the k lowest velocities resulting in CPs. Figure 3 shows a notional example of the up-down method. For this example, the initial estimate is 2500 ft/s with $\delta = 50$ ft/s. The estimate for V_{50} is 2495 ft/s using $k = 5$, but note that separation occurs in this test sequence because there is no overlap in responses. As a result, the response curve cannot be estimated.

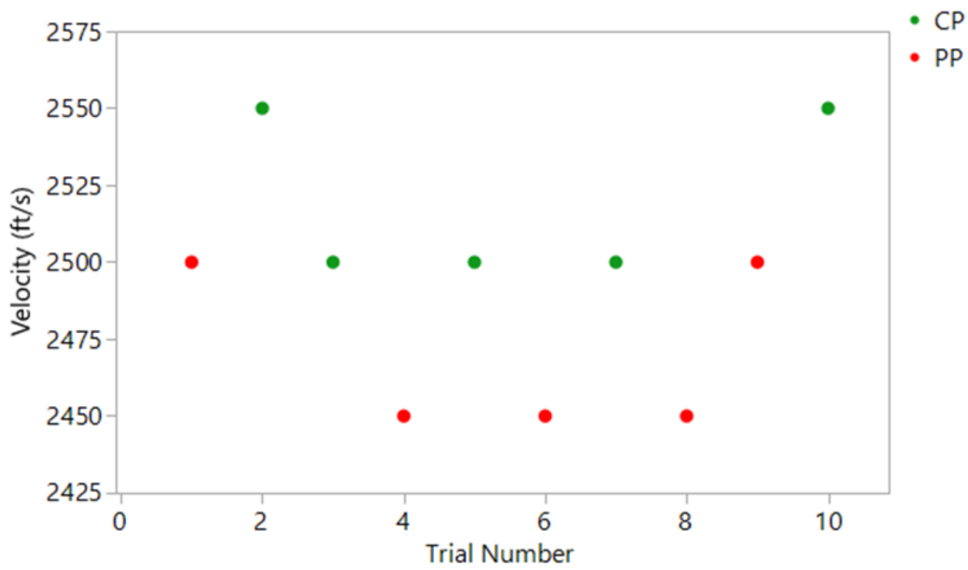

Figure 3. Notional example of up-down method

The up-down method was designed to converge at V_{50} and thus an average of these velocities can provide a reasonable estimate of V_{50}. The up-down method is very easy to implement as it does not require special statistical software or difficult calculations, but it also has several drawbacks. This approach is designed to only estimate one point on the response curve (V_{50}) and typically cannot estimate the entire response curve due to separation occurring. It is also not an effective method to estimate other velocities, particularly at the extremes (e.g., V_{10}). The method is also highly dependent on the initial starting value. Without historical information to estimate μ_G, the method will take longer to converge to an estimate. Finally, because there is a constant step size (δ), the method may not converge to V_{50} in some cases (choosing a step size that is too large).

Langlie Method

Another common DoD method is the Langlie method (Langlie, 1962), a sequential experiment that does not require an initial estimate of V_{50}. This method was developed for use in DoD testing for experiments limited to 15 to 20 runs. A modified version of the Langlie method is implemented in the Army (Collins and Moss, 2011). A lower and upper limit (μ_{min} and μ_{max}) are initially selected as the test interval. This interval should be made sufficiently wide since it serves as the range of stress levels tested. The subsequent test points, a modified version of the Langlie method, are chosen as follows (Collins and Moss, 2011):

1. $x_1 = (\mu_{min} + \mu_{max})/2$
2. If y_1 is a CP, $x_2 = (x_1 + \mu_{min})/2$; if y_1 is a PP, $x_2 = (x_1 + \mu_{max})/2$.
3. If y_1 is a CP and y_2 is a PP or y_1 is a PP and y_2 is a CP, $x_3 = (x_1 + x_2)/2$.
   If y_1 and y_2 are both PPs, $x_3 = (x_2 + \mu_{max})/2$.
   If y_1 and y_2 are both CP, $x_3 = (x_2 + \mu_{min})/2$.
4. If the previous shots are all CPs, adjust the lower and upper limits such that $\mu_{min} = \mu_{min} - 65$ ft/s and $\mu_{max} = \mu_{max} - 65$ ft/s. Then $x_4 = (x_3 + \mu_{min})/2$.*
   If the previous shots are all PPs, adjust the lower and upper limits such that $\mu_{min} = \mu_{min} + 65$ ft/s and $\mu_{max} = \mu_{max} + 65$ ft/s. Then $x_4 = (x_3 + \mu_{max})/2$.*
5. In general, x_{i+1} is determined after completing i trials by working backward in the test sequence. Beginning at the ith trial, work backward until there is a trial p such that there are an equal number of CPs and PPs in the pth through ith trials. Then $x_{i+1} = (x_i + x_p)/2$.
   If there is no value p such that there are an equal number of CPs and PPs before the ith trial, $x_{i+1} = (x_i + \mu_{min})/2$ or $x_{i+1} = (x_i + \mu_{max})/2$ if y_i is CP or PP, respectively.

*The manual for the Langlie method defines this adjustment as 20 m/s, which we have translated to ft/s for consistency.

The method continues for at least 8 trials and stops when all stopping criteria are met. The stopping criteria for this modified method are:

1) At least one PP has a higher velocity than a CP (i.e. there are overlapping responses and separation is broken)
2) The average velocity of CPs is larger than the average velocity of PPs
3) The range of the tightest three PPs and three CPs is within 125 ft/s
4) There exist test points approximately ± 65 ft/s from V_{50} estimated from the tightest three PPs and three CPs.

Figure 4 shows a notional test sequence using the Langlie method where $\mu_{min} = 2400$ ft/s and $\mu_{max} = 2700$ ft/s. Using all test points, a statistical software package estimates the logistic regression model for CP as:

$$\widehat{\mathrm{P}}(CP|velocity) = \frac{1}{1 + e^{-(-465.77 + 0.18 \times velocity)}} \quad (2)$$

Using this model, V_{50} is estimated as 2551.9 ft/s with an 80% CI of (2532.95, 2569.00). V_{10} is estimated as 2538.86 ft/s with an 80% CI of (2397.50, 2547.90).

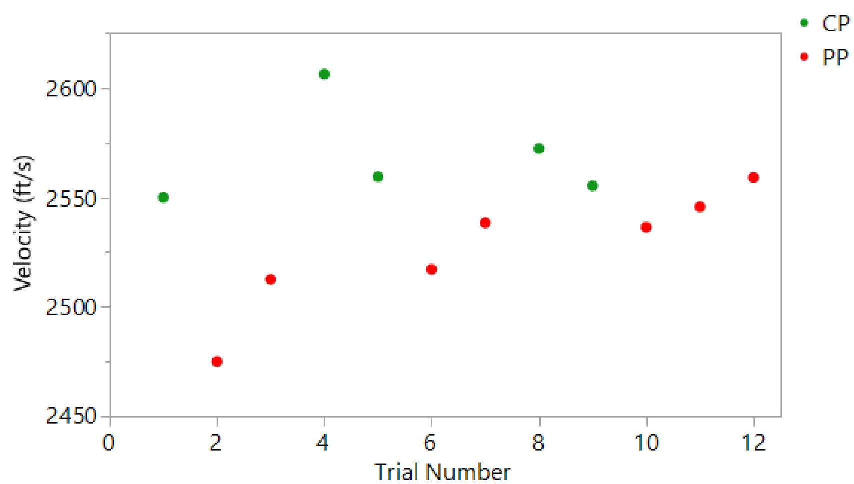

Figure 4. Notional example of the Langlie method

The Langlie method was developed to estimate V_{50}; however, because one of the stopping criteria for the modified method requires breaking separation, the resulting data can be used to estimate the entire ballistic response curve. Unlike the up-down method, the step size at each step is not constant and depends on the results of the previous shots. Another advantage of the Langlie method is that the only required inputs are a minimum and maximum velocity. If quantiles other than V_{50} are of interest in the experiment, however, the Langlie method will not perform as efficiently as it does for estimating V_{50}. For example, Figure 5 shows the response curve generated from the test sequence in Figure 4 with associated 80% CIs for various quantiles. While the CI for V_{50} is relatively narrow, the intervals for the other quantiles are imprecise. Estimates of V_{10} tend to be biased toward V_{50} because more test points are placed near V_{50} by design. As we will demonstrate later, this method is not as efficient or as robust as 3pod.

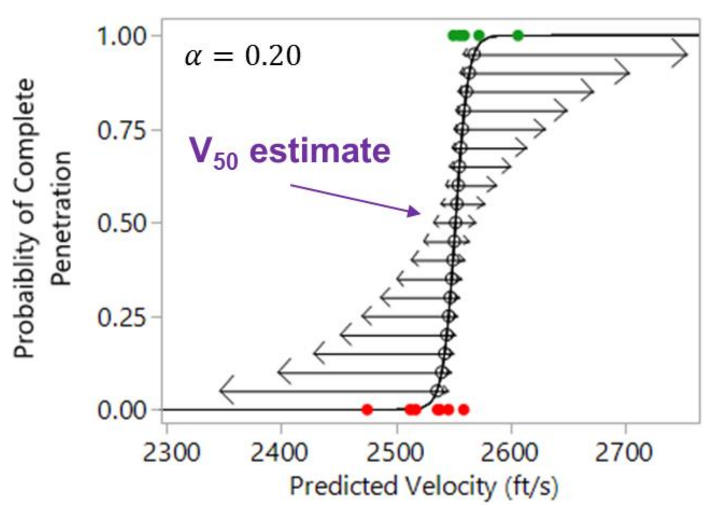

Figure 5. Response curve with 80% CIs for Langlie example test sequence

Three-Phase Optimal Design

Because of the limitations of the methods previously described, an alternative approach that can be performed successfully in a relatively small number of runs and can estimate the entire response curve is preferred. 3pod is a sequential test strategy developed by Wu and Tian (2013) for sensitivity experiments (which include ballistic resistance testing). The 3pod method has three sequential steps that can be summarized as Search, Estimate, and Approximate. In the "Search" phase, the goal is to identify a reasonable experimental range of velocity and to bound the ballistic response curve with CPs and PPs. This phase has three steps:

I1: Obtain at least one CP and one PP to bound the range of velocities

I2: Search for overlapping regions so that the largest velocity among the PPs (M_0) is greater than the lowest velocity among the CPs (m_1) (i.e. break separation)

I3: Enhance this overlapping region using one or two shots, particularly when the difference between M_0 and m_1 is small

Phase 1 (shown in the flowchart in Figure 6) finds the overlapping region of CPs and PPs in order to estimate the parameters in the resulting ballistic response curve. The second phase, "Estimate," seeks to optimize the model parameters of the resulting model of the ballistic response curve. Test points in phase two are chosen sequentially to optimize a measure of the test matrix (the D-optimality criterion, see Wu and Tian [2013] for details) that includes all previous test points. The user specifies the number of test points in this stage. Finally, the "Approximate" phase, an optional phase, is used to better approximate a specific point on the response curve (such as V_{10} or V_{50}). The number of test points in this phase is also user-specified. Velocities are sequentially chosen in this phase using an approximation procedure called Robbins-Monro-Joseph (Wang et al., 2015). This final phase can be skipped in order to better estimate the overall response curve, particularly when there are a limited number of runs available in the experiment. The resulting test points from the three phases can be analyzed using a logistic regression model. The estimated response curve from this analysis can then be used to estimate any quantile value of the response curve such as V_{10} or V_{50} with CIs. To begin the method, the user must specify a minimum and maximum velocity as well as an estimate of the standard deviation of velocity.

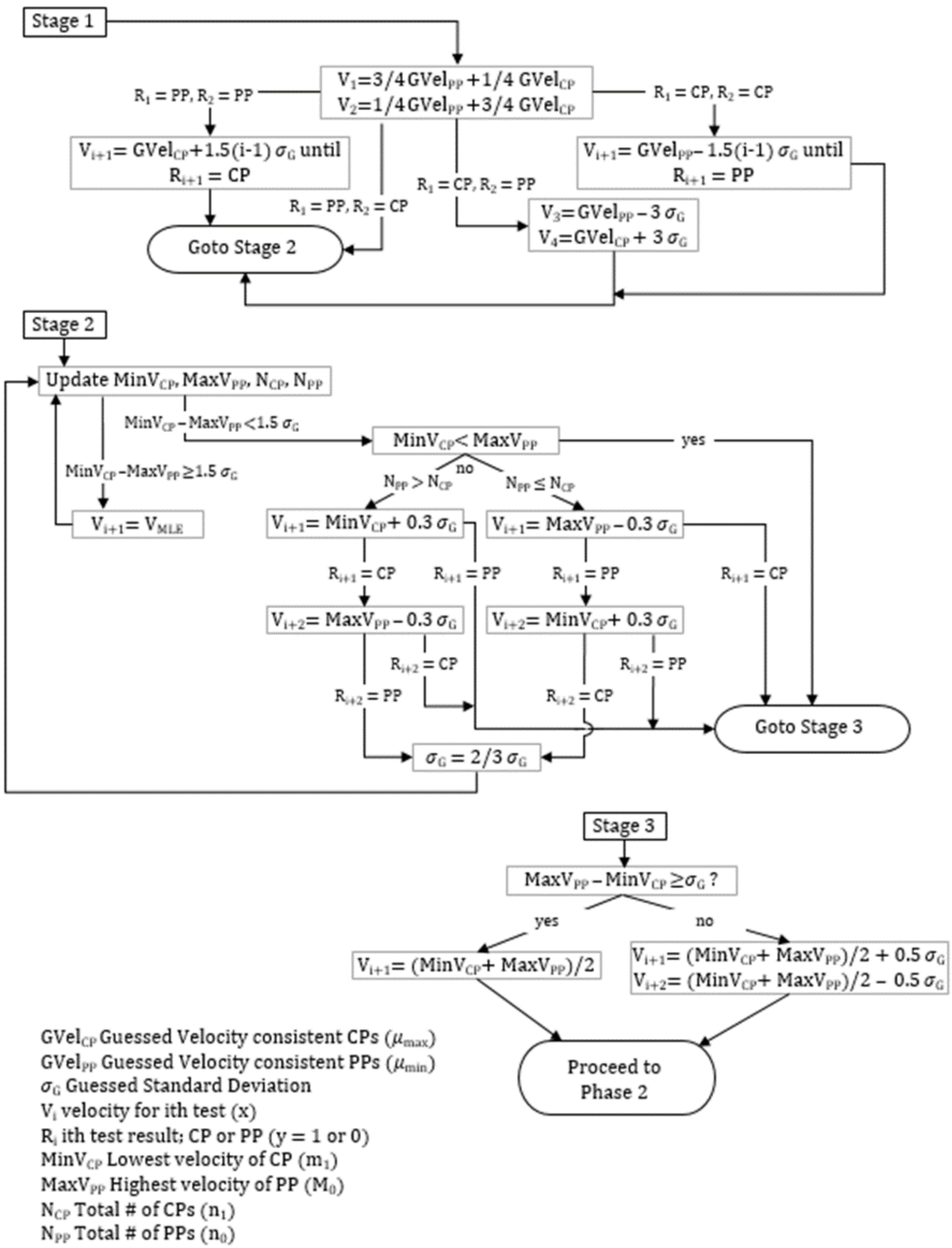

Figure 6. Flowchart of phase 1 of the 3pod method

An Illustrative Example

Suppose a program needs to estimate the velocity at which there is a 10% chance of CP. Previous testing indicates the velocities for PPs and CPs range from 2400 to 2700 ft/s. Due to limited resources, only 28 runs are available in the test. To initiate the test, the required inputs are a minimum velocity (2400 ft/s), maximum velocity (2700 ft/s), and estimated standard deviation of velocity (50 ft/s). Figure 7 shows a simulated test sequence for this notional ballistic resistance test. Eight runs were required to complete phase 1; the test team evenly splits the remaining 20 runs between phase 2 and phase 3 so that the complete test has 8 runs in phase 1, 10 in phase 2, and 10 in phase 3. Phase 1 is shown broken into the three steps outlined previously (I1, I2, and I3). Note that the velocity of the sixth shot (which is a PP) is larger than the velocity of the fourth shot (which is a CP). This means that there is no separation between the responses (i.e., there is overlap in velocities of the CPs and PPs). This overlap allows us to fit a logistic regression model and estimate any velocity V_p. The sixth shot triggers the final step of phase 1 (shots 7 and 8). Phase 2 occurs in shots 9 through 18, where each velocity tested is determined via an optimality criterion. Finally, phase 3 shows small, incremental changes in the velocity. The procedure in this phase is designed to approximate a point along the curve as specified by the user; in this case, V_{10}.

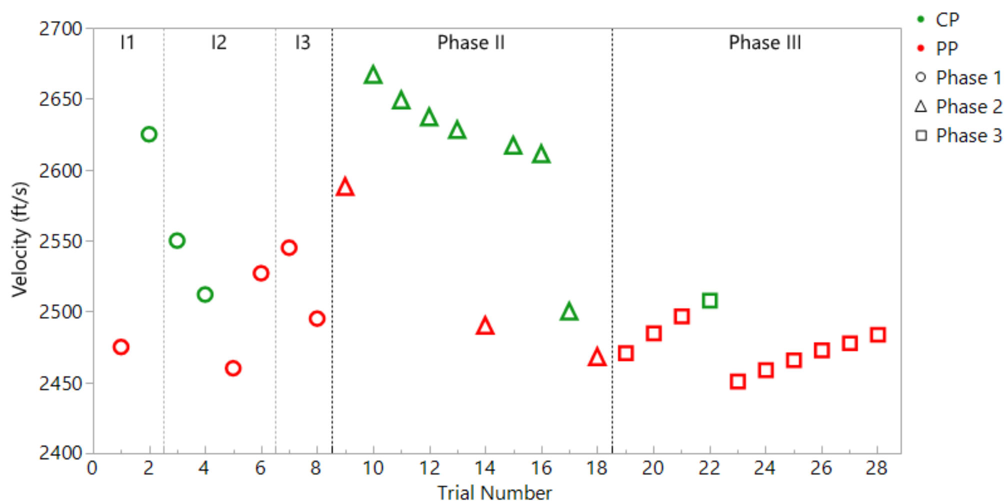

Figure 7. Sample test sequence for a ballistic resistance test. Phase 3 is designed to approximate V_{10}

Figure 8 shows the estimated ballistic response curve using all 28 runs from the experiment with the results of the test overlaid on the plot. The test points across the top of the figure resulted in CPs while those on the bottom resulted in PPs. The curve was fit using a logistic regression model so that the probability of CP is estimated as

$$\hat{P}(\text{CP}|velocity) = \frac{1}{1 + e^{-(-87.08+0.034*velocity)}} \quad (3)$$

The parameters for the model were estimated using statistical software. Figure 8 also shows 80% CIs for the estimated velocity at V_{05}, V_{10}, V_{15}, etc. The estimate of V_{10} in this example is 2478.2 ft/s with an 80% CI of (2431.0, 2502.3) ft/s. The estimate of V_{50} is 2542.4 ft/s with an 80% CI of (2520.0, 2576.3) ft/s.

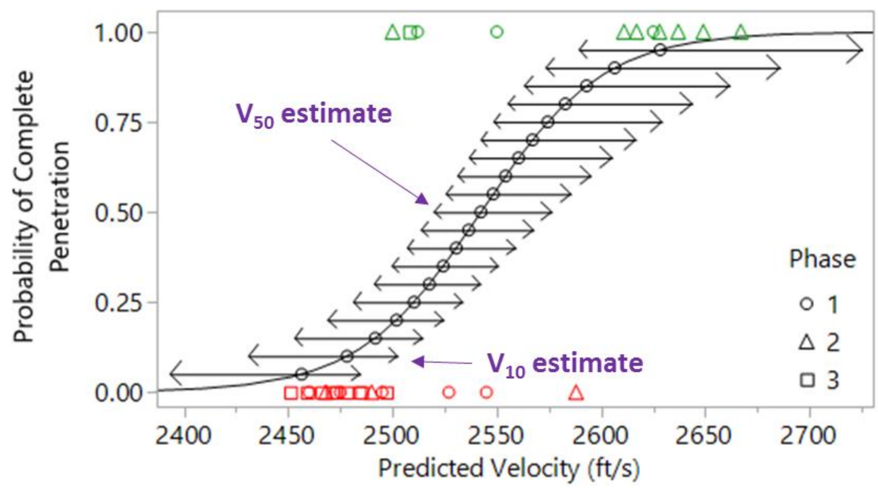

Figure 8. Estimate response curve using all 28 runs in the 3pod experiment

Number of Runs in Phase 1 of 3pod

One question to explore further is the number of runs required to complete phase 1 in 3pod. Completing phase 1 is essential in order to estimate the ballistic response curve. However, the number of runs required in phase 1 is not a fixed number, which can lead to uncertainty in test planning. We performed a simulation study under a variety of conditions for the armor test example discussed previously. The starting minimum and maximum velocities remain the same throughout each scenario ($x_{min} = 2400$ and $x_{max} = 2700$) with the guessed mean μ_G defined as $(x_{min} + x_{max})/2 = 2550$ ft/s. The guessed standard deviation σ_G was chosen to be either 25 or 50 ft/s. The true mean μ_T was varied between three conditions: $\mu_T = \mu_G$ $(\Delta\mu = 0)$, $\mu_T = \mu_G + 100$ $(\Delta\mu = 100)$, and $\mu_T = \mu_G - 100$ $(\Delta\mu = -100)$. The true standard deviation σ_T was also varied between three conditions: $\sigma_T = \sigma_G$ $(\Delta\sigma = 0)$, $\sigma_T = \sigma_G + 10$ $(\Delta\sigma = 10)$, and $\sigma_T = \sigma_G - 10$ $(\Delta\sigma = -10)$. A factorial of these combinations was implemented with each of the 18 scenarios simulated 10000 times. The results of this simulation study are summarized in Figure 9. The median number of runs to complete phase 1 of 3pod ranged between 8 and 12 and the 75th percentile was always 14 or less across all 18 scenarios. Note that the number of runs required tends to be higher for a smaller standard deviation.

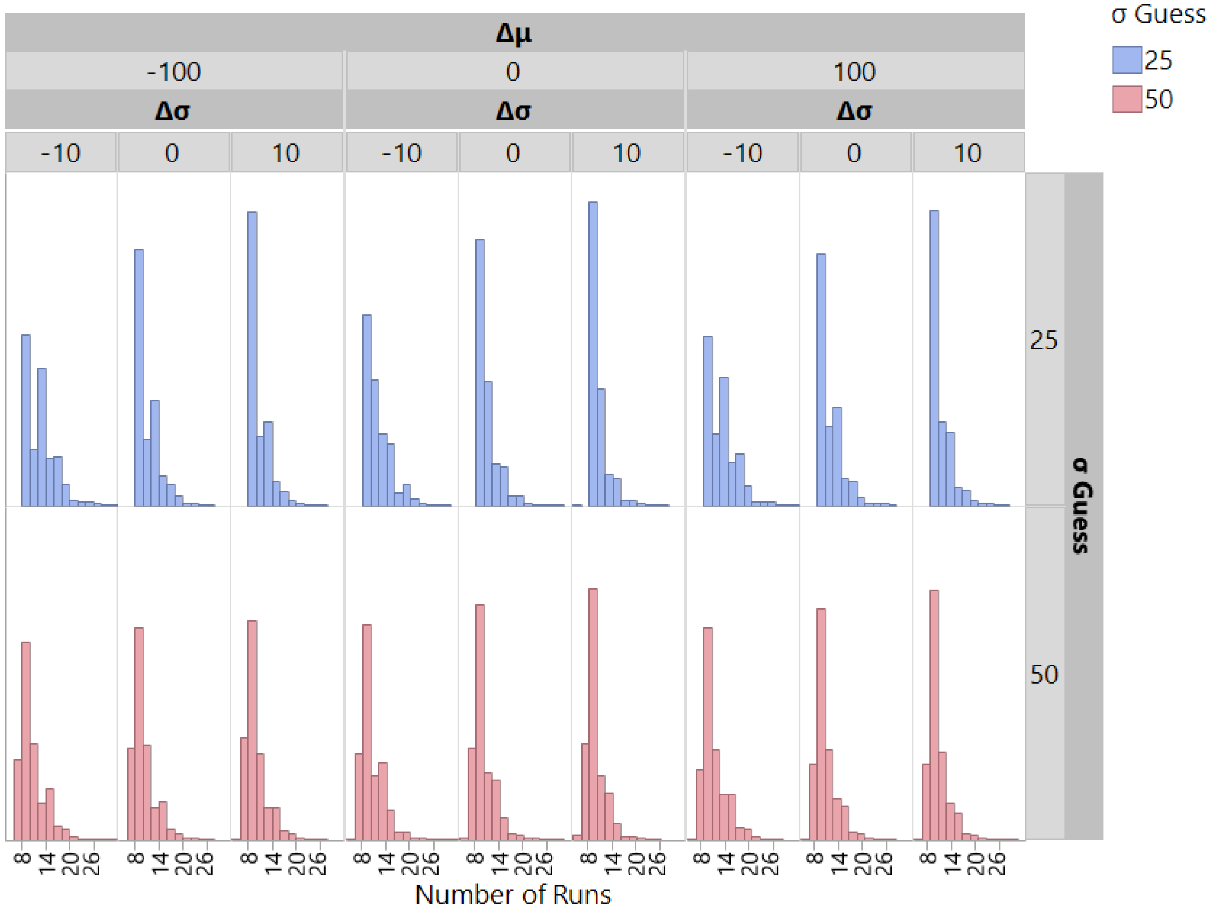

Figure 9. Summary of Simulation Results

How to Implement 3pod

An application implementing 3pod is available on the Test Science website ("3Pod") or the shinayapps.io website (tomsshinyapps) (shown in Figure 10 was built using the statistical software package R, but does not require the user to have R to use the app. To begin, the user enters the initial settings for the experiment as described previously, the desired significance level (1 – confidence level) for the resulting CIs, the total number of runs, and the proportion of remaining runs after phase 1 is complete to allocate to phase 2. For example, to split the remaining runs evenly between phases 2 and 3, the user should select 0.5 for this option.

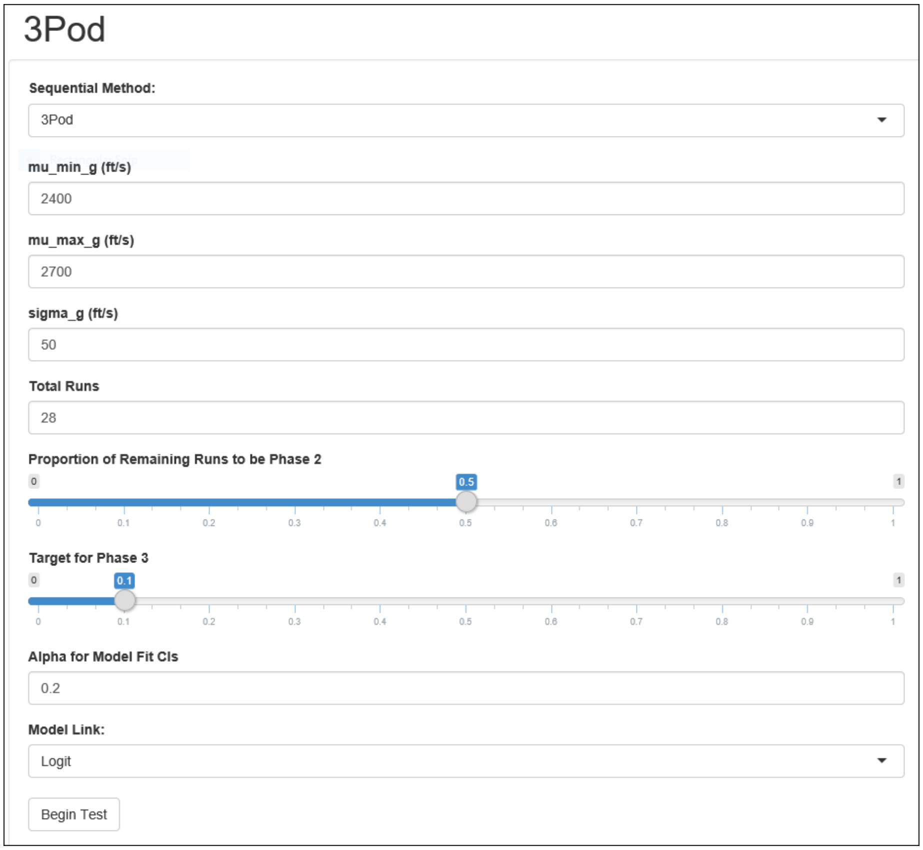

Figure 10. Entering initial algorithm parameter in the 3pod App

Once all initial settings have been entered, the app displays the first velocity at which to shoot. After the results of the first shot are entered into the app (CP or PP and actual velocity achieved) as shown in Figure 11, the next test point (velocity) is displayed. Using these results, the algorithm determines the velocity setting for the next run using the procedure described earlier. Plots similar to those in Figure 7 and Figure 8 are produced by the app. The results table can be downloaded for further analysis at the end of the test. Source code in R developed by The Armament Research Development and Engineering Center (ARDEC) Statistics Group at Picatinny Arsenal is also available. An R package implementing that code is currently under development.

Figure 11. Using the 3pod App

Additional Study of 3pod

A simulation study performed by research analysts at the Institute for Defense Analyses (IDA) (Johnson et al., 2014) compared several sequential testing strategies for sensitivity experiments, including the up-down method, Langlie method, 3pod, and others under a variety of conditions. The velocity is assumed to have a "true mean" μ_T and "true standard deviation" σ_T. Each scenario in the simulation considers the effect of choosing a guessed mean μ_G and guessed standard deviation σ_G to select the initial conditions of each test strategy. The step size used in the up-down method is σ_G. The starting values in the Langlie method and 3pod are $x_{min} = \mu_G - 4\sigma_G$ and $x_{max} = \mu_G + 4\sigma_G$. We summarize a few general results here and refer the reader to Johnson et al. (2014) for complete details. Unsurprisingly, all three methods perform similarly when the true mean and the guessed mean are the same. The median error of V_{50} is centered near 0 for all three methods in this scenario.

In general, 3pod was shown to be the most robust method compared to the other three over all scenarios. In other words, the estimates of V_{50} and V_{10}, both in terms of bias and spread, were not largely affected by using μ_G and σ_G that varied from the true values. The up-down method does not provide accurate estimates of V_{50} when the guessed initial conditions are different than the true values. Among these three methods, 3pod performs best for estimating V_{10}, followed by the Langlie method. The up-down method was the worst compared to all the methods in estimating V_{10}. 3pod tended to decrease the median V_{10} error while the Langlie and up-down methods increased the median V_{10} error. In general, the study indicated 3pod was the most robust method in estimating multiple quantiles. While the Langlie method did not perform as poorly as the up-down method, it was inferior to 3pod in most instances.

Conclusion

We presented several test strategies for an experiment with one stress factor and a binary response. While we demonstrated the methods for a ballistic resistance test, the methods are applicable in other similar experiments. Compared to LASP, the up-down method, and the Langlie method, 3pod provides more information with the same test resources where there is one factor of interest and the response is binary. We encourage users to consider using 3pod in these types of experiments to obtain efficient estimates of the response curve. For additional assistance, contact the STAT COE at COE@afit.edu.

References

"3Pod", *Test Science*, testscience.org/design-execute/constructing-test-designs/3pod/3pod/

Collins, J.C. & Moss, L.L.C., "LangMod Users Manual", Report No. ARL-TN-37, *US Army Research Lab*. June 2011, www.arl.army.mil/arlreports/2011/technical-report.cfm?id=6192

Harman, Michael. "Use of the Binomial Nomograph for Test and Evaluation Planning." Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 2013.

Johnson, T. H., Freeman, L., Hester, J., & Bell, J. L. (2014). A Comparison of Ballistic Resistance Testing Techniques in the Department of Defense. *IEEE Access*, 2, 1442-1455.

Langlie, H.J. (1962). A Reliability Test Method for "One-Shot" Items. Aeronutronic Division of Ford Motor Company Publication No. U-1792.

"MIL-STD-662F", *Defense Logistics Agency (DLA)*, 18 December 1997, quicksearch.dla.mil/qsDocDetails.aspx?ident_number=35877

Myers, Raymond H., et al. *Generalized Linear Models with Applications in Engineering and the Sciences*. Wiley, 2010.

Tomsshinyapps, "3Pod", *shinyapps.io*, tomsshinyapps.shinyapps.io/3pod/

Truett, Lenny. "Using Operating Characteristic Curves to Balance Cost and Risk-Best Practice." Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 2013.

Wang, D., Tian, Y., and Wu, C.F.J. (2015). A Skewed version of the Robbins-Monro-Joseph procedure for binary response. *Statistica Sinica*, 25, 1679-1689.

Wu, C.F.J., Tian, Y. (2013). Three-phase optimal design of sensitivity experiments. *Journal of Statistical Planning and Inference*, 149, 1-15.