AFRL-RW-EG-TR-2009-7005

ANALYSIS OF SENSITIVITY EXPERIMENTS – A PRIMER

Douglas V. Nance

Air Force Research Laboratory
Munitions Directorate
AFRL/RWAC
101 W. Eglin Blvd.
Eglin AFB FL 32542-6810

NOVEMBER 2008

TECHNICAL REPORT FOR PERIOD AUGUST 2008 - NOVEMBER 2008

AIR FORCE RESEARCH LABORATORY, MUNITIONS DIRECTORATE

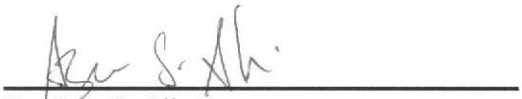
■ **Air Force Materiel Command** ■ **United States Air Force** ■ **Eglin Air Force Base**

NOTICE AND SIGNATURE PAGE

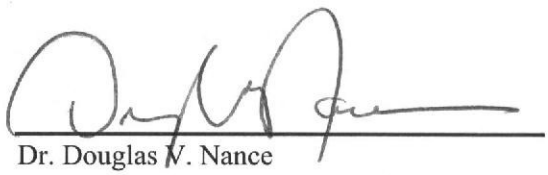
.

AFRL-RW-EG-TR-2009-7005 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:



Dr. Azar S. Ali
Technical Advisor
Assessment and Demonstration Division



Dr. Douglas V. Nance
Senior Engineer
Computational Mechanics Branch

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
15-01-2009	TECHNICAL	August 2008 – November 2008

4. TITLE AND SUBTITLE
ANALYSIS OF SENSITIVITY EXPERIMENTS - A PRIMER

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62602F

6. AUTHOR(S)
Douglas V. Nance

5d. PROJECT NUMBER
2502

5e. TASK NUMBER
67

5f. WORK UNIT NUMBER
63

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Munitions Directorate
AFRL/RWAC
101 W. Eglin Blvd.
Eglin AFB, FL 32542-6810

8. PERFORMING ORGANIZATION REPORT NUMBER
AFRL-RW-EG-TR-2009-7005

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Munitions Directorate
AFRL/RWMI
101 W. Eglin Blvd.
Eglin AFB, FL 32542-6810

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL-RW-EG

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RW-EG-TR-2009-7005

12. DISTRIBUTION / AVAILABILITY STATEMENT
DISTRIBUTION A: Approved for public release; distribution unlimited. 96th ABW/PA Approval and Clearance # 96ABW-2008-0024, dated 19 Dec 2008.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report is expository in nature and is intended to explain the basic concept of a sensitivity experiment as well as the analysis of its data. These experiments are widely applicable to engineering problems that involve binary (pass/fail of go/no-go) outcomes. We begin by introducing the simple idea behind a sensitivity experiment; then we describe the basic 'Up and Down' testing method. Particular emphasis is placed upon the parameter this procedure is intended to identify. Next, we describe Garwood's method, a type of Probit analysis, for analyzing sensitivity test data. A specialized version of this scheme is derived for stable digital computation. Confidence interval estimation is discussed along with an analysis of variance. A set of example problems are solved; our results are compared with archival solutions.

15. SUBJECT TERMS
sensitivity tests, probit, normal distribution, detonation, reliability

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Douglas V. Nance
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL	54	**19b. TELEPHONE NUMBER**

ANALYSIS OF SENSITIVITY EXPERIMENTS – A PRIMER

by

Douglas V. Nance
AFRL/RWAC
101 W. Eglin Blvd
Eglin AFB, FL 32542-6810

January 2009

ABSTRACT

This report is expository in nature and is intended to explain the basic concept of a sensitivity experiment as well as the analysis of its data. These experiments are widely applicable to engineering problems that involve binary (pass/fail of go/no-go) outcomes. We begin by introducing the simple idea behind a sensitivity experiment; then we describe the basic “Up and Down” testing method. Particular emphasis is placed upon the parameter this procedure is intended to identify. Next, we describe Garwood’s method, a type of Probit analysis, for analyzing sensitivity test data. A specialized version of this scheme is derived for stable digital computation. Confidence interval estimation is discussed along with an analysis of variance. A set of example problems are solved; our results are compared with archival solutions.

PURPOSE

The purpose of this report is to document in-house research concerned with the analysis of sensitivity test data. The technical information developed in the course of this effort is of significant importance in enabling oversight for go/no-go testing efforts conducted within the Department of Defense.

OBJECTIVES

We begin with a basic discussion of sensitivity experiments based upon a common example accessible to the layman. In this example, we emphasize the statistic (the number or numbers that we hope to extract from the data and their meaning(s)) of primary interest to the analyst, and we also describe how a simple series of tests is conducted to obtain the data. The second section of the report derives the analytical method used to process the test data and extract the pertinent statistics of interest. Later, we derive equations for the calculation of confidence intervals and analyzing the variance of parameters. Finally, we present results for a set of example problems.

TABLE OF CONTENTS

1 INTRODUCTION

Statistics is a complicated science. Having terrorized dormitories and classrooms across the country, its late night, torturous study sessions among servile students have earned it a dubious “GPA-busting” reputation and the forlorn appellation *Sadistics*. Although it is a division of mathematics, it stands aside from the mainstream of this discipline. In fact, its more skillful practitioners refer to themselves as *probabilists* instead of mathematicians. Statistics can also incite chagrin among practicing engineers and scientists. In a recent meeting with the author, this assertion was reaffirmed as one professional uttered the phrase, “*I hate statistics!*”

Why does statistics evoke such negative emotion? There is a reasonable answer for this question. Although many professionals are required to use it, they lack the theoretical background and experience required in order to understand how statistics works. It is a branch of mathematics that differs from subjects like calculus or linear algebra. In these subjects, an individual can learn a few basic ideas and apply them in a nearly “cookbook” way to solve problems. Statistics is not so cooperative. Statistical or stochastic theory is deeply buried in advanced mathematical theory involving topics such as *Lesbegue Integration, Measure Theory* and *Functional Analysis*. If these topics are well understood, then probabilistic theory is more accessible, and statistics becomes understandable. Unfortunately, few people have the time and patience required to study and master these disciplines. Hence, statistical theory remains arcane, and its mastery eludes all but its diehard practitioners.

The difficulty associated with mastering statistical inference presents a true dilemma. Statistics is an extremely applied science. It has an extensive number of

applications in practically every field of engineering or scientific endeavor. But in order to obtain good computational results, great care is required while converting abstract statistical theories into practice. After having completed graduate study in stochastic processes, the author now agrees, at least in part, with his venerable instructor and thesis supervisor. Statistics is best taught and understood through the use of clearly explained examples along with carefully planned excursions into probability theory. We have attempted to take this approach in the discussions that follow. He who dares to venture directly into the world of stochastic theory is doomed to become lost, perhaps forever.

This report focuses on analytical methods used to process binary and binomial statistical trials. A single binary trial or experiment has only two possible outcomes, either a *success* or a *failure*. A binomial trial may be thought of as a series of binary trials taken at the same "level". The result of a binomial trial consisting of n binary experiments is say, p successes and $(n - p)$ failures with a crude probability of success of p/n for a particular binomial trial. A binary trial is a special case; it is a binomial trial with n = 1 and p can be either 0 or 1. As it happens, *Sensitivity Tests* are characterized by a mixture or series of binary and binomial trials. For military applications, these tests have a great deal of utility since many items of military hardware may be used only once and may only be evaluated by pass/fail criteria. Many munitions can be thought of in this way. For this reason, we describe the basic features of a sensitivity test in the next section of this report.

1.1 Sensitivity Tests

A general sensitivity test consists of a finite series of either binary or binomial experiments (or a mixture of the two). Each experiment is performed at a *level* defined by distinct value of an explanatory variable.[1] The value of the explanatory variable directly corresponds to the measurable *dosage* for this level. The dosage is the related to the *stimulus*, an unmeasurable quantity that drives the outcome of the experiment (or trial) at the chosen level.[2] The stimulus may be thought of as a mechanistic (physical, biological, etc.) process that results from the dosage. Each binary trial has only two possible outcomes, a success or a failure. If a binomial trial (consisting of n binary trials) is conducted at a given level, then without any loss of generality, we can say that p *successes* and $(n - p)$ *failures* will result at this level. A basic assumption behind the sensitivity test is that as the dosage (stimulus) increases, then the probability of success also increases. That is to say, at "lower" levels of stimulus, we expect more failed trials. As the stimulus increases at higher levels, we expect more successful trials, up to the point where practical all individual binary trials are successes (or for binomial trials, p equals n).[3] Unfortunately, due to the pass/fail nature of each trial's outcome, we cannot *precisely* determine the dosage that will result in a success. Instead, all that we can do is select a dosage and via test determine whether the *critical dosage* is higher or lower than the test dosage.[4] That is to say, if the test is a success, the critical dosage is less than or equal to the test dosage. Conversely, if the test is a failure, then the critical dosage is greater than the test dosage. To promote greater understanding, let us consider a basic example of sensitivity testing.

Suppose that a pharmaceutical manufacturer has developed a new antibiotic to fight a particular strain of bacterial pneumonia. As a normal part of the certification procedure, an effective dosage must be estimated for this drug. *The critical dosage* is defined as the mass (say, in milligrams) of the drug that must be administered to eradicate all invading bacteria *with no excess antibiotic remaining in the body*. As you may imagine, this dosage is impossible to measure. Also, the outcome of the test is determined by the stimulus. The stimulus for the problem requires a detailed knowledge at the microscopic scale of how the antibiotic interacts with each individual bacterium. This information is not known and cannot be determined, so we must obtain an estimate of the effective dosage by sensitivity testing.

To estimate the effective antibiotic dosage, we select a set of dosage levels for testing, for example {0, 10, 20, 30, 40, 50} milligrams. Naturally, this set defines six dosage levels for the sensitivity test. We then select a number of healthy test animals (say, Rhesus monkeys) for use and decide how many are to be tested. In as much as is possible, the test animals (or test articles) should be chosen so that they have the same anatomical and physiological (or physical) characteristics.[5] If so desired, we can formulate the entire sensitivity test series as a sequence of binary trials, one test animal per trial. To conduct a trial in this sequence, we choose a level and infect a test animal with the pneumonia bacteria. Then we wait a pre-specified amount of time and administer the dosage for the chosen level to the test animal. In the period of time after administering the drug, we determine the response; the test animal either lives (a success) or dies (a failure). The *response* data is collected, and we then move onto the

next level and continue testing. At the conclusion of the entire sequence, we have determined a number of successes and failures at each dosage level.

This example brings an important fact to light. As we move between trials, we cannot reuse test articles, (e.g., test animals). Why? Obviously, if a test animal dies at the preceding trial, it cannot be reinfected and tested. If the animal survives the preceding trial, its physiology has been altered by the presence of the antibiotic and by the action of its immune system. The animal's bloodstream, lymphatic system and pleural tissues are inundated with antibodies to the pneumonia. As a result, it cannot be equivalently reinfected, so the results of a subsequent test would be tainted. Once a test article has been exposed to the stimulus, it should not be tested again.

The output of a complete sensitivity test series may be represented as a plot of percentage success (*p*/*n*) versus the dosage (or explanatory variable) level. We observe that the change in the success percentage is smaller for low and high dosages while the greatest change occurs near the mean dosage. Hence, we assume that the dosage-response curve follows the cumulative normal distribution function.[5] When a sufficient amount of data has been collected, this assumption is testable.[5] With an appropriate number of tests, we can determine the "50% point" for the distribution, the level of dosage that causes a successful response for 50% of all like test articles.[2] This value of dosage is the *mean* for the distribution. We are also interested in how success percentages behave for dosages further away from the mean. The property is termed *dispersion* and is measured by the *standard deviation*. The primary focus of post-test analysis resides in determining these properties for the distribution.

1.2 The Bruceton Testing Method

The mean μ and standard deviation σ must be estimated accurately if our analysis is to have real meaning. For this reason, the sensitivity test procedure is designed to concentrate measurements around the mean in order to refine the estimate. The Bruceton or "Up and Down" test is commonly used for this purpose.[4] The first step in conducting a Bruceton test is to decide the range of dosages (minimum to maximum) as well as the dosage levels. We must also decide the total number of individual binary trials to be performed during the test series. In practice, we require a minimum of 20 trials, but twice that number is recommended since under theoretical limitations, the effective sample size is usually half of the actual sample size.[4] Table 1 illustrates a notional Bruceton test series. The results of the initial trial are recorded in column 2 while the response for the last trial is recorded in column 21. For the initial trial, we choose a dosage (level 2 in the example) that is believed to be closest to the actual mean. If the outcome of this trial is a success, we conduct the next trial at the dosage just below the initial level. On the other hand, if the first trial is a failure, we conduct the second test at the dosage level just above the initial level. The dosage level for the third trial is determined in the same manner but is based upon the outcome of the second trial. Dosage levels for the remaining trials are determined in the same way, but in most cases, we do not use all of the data collected. The reason for excluding some of the data is based upon our desire to resolve the normal mean.

As we stated above, Bruceton testing concentrates measurements in the vicinity of the mean dosage. Since the change in dosage level between trials "opposes" the

Table 1. Illustration of responses for a Bruceton test series consisting of 20 binary trials. The notation "x" denotes a success while "o" denotes a "failure". Six dosage levels are numbered 1 to 6 from low to high.

C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 10	C 11	C 12	C 13	C 14	C 15	C 16	C 17	C 18	C 19	C 20	C 21
Lev 1						o														
Lev 2	o				x		o						o							
Lev 3		o		x				o		o		x		o						
Lev 4			x						x		x				o		o			
Lev 5																x		o		o
Lev 6																			x	

Table 2. Summary of the Bruceton test results extracted from Table 1 shown for each dosage level.

Dosage Level	1	2	3	4	5	6
Successes (x)	0	1	2	3	1	1
Failures (o)	1	3	4	2	2	0
Total No.Trials	1	4	6	5	3	1

sense of the preceding response, the up and down nature of the test series attempts to center data collection around the mean. Unfortunately, if our initial guess for the mean dosage level is poor, we automatically introduce extraneous data into the test series. Suppose that we have guessed an initial value for the dosage that is too low. Then for two or more tests, we will steadily increase the dosage level and obtain failed responses. For example, see levels 2 and 3 in Table 1. Let us assume that we obtain a success on the third trial (shown in column 4). This sudden change in the response (indicated by the red block in Table 1) is termed as a *reversal*. It follows that we avoid introducing extraneous “start-up” data by analyzing only information obtained beginning with the second trial. In this way, data collection tends to remain near the mean. The results of this example test series are shown versus dosage level in Table 2.

1.3 Difficulties with Data Reduction

Sensitivity testing does have its caveats. In the first place, the test data must contain a *zone of mixed results*.[6] That is to say, the highest level at which a trial fails must exceed the lowest level observed for a successful trial. If this condition is not satisfied, the equations for estimating σ become inconsistent, and zero is the only value that can be obtained for σ.[7] In the example shown in Table 1, "2" is the lowest level corresponding to a success while "5" is the highest level containing a failure. Hence, for this example, we obtain a zone of mixed results between levels "2" and "5". The Bruceton test procedure is usually successful in establishing the mixed zone.[6]

Sensitivity testing also possesses certain intrinsic weaknesses. Since it concentrates on resolving the distribution mean, it loses accuracy near the "tails" of the distribution, i.e., the regions near 0% and 100% response.[4] For reliability problems, we are most interested in the region corresponding to response values exceeding 99%. As a result, care must be taken when sensitivity tests are conducted with this purpose in mind. Due diligence must be paid to the structure of the dosage levels and to the number of trials. The chosen data analysis methodology is equally important since the choice of distribution can have a significant effect in the high probability region. In certain cases, the logistic distribution is chosen over the normal distribution due to its more conservative nature in the "tails". That is to say, the logistic distribution tends to report slightly lower cumulative response than does the normal distribution.[8] This shortcoming emphasizes the importance of correctly calculating confidence intervals for system reliability probabilities in the tail regions. Secondly, the maximum likelihood

estimation (MLE) procedure used to analyze the test data generally requires a large data set. For small data sets, MLE loses "efficiency"; the estimates of μ and σ become biased and acquire excessive variance.[7] Moreover, MLE may not even be able to analyze certain small data sets. A classic example is that of a test series without a mixed zone of results.[7] Note that Dixon and Mood have published the disclaimer: "Measures of reliability may be very misleading if the sample size is less than forty or fifty."[4] To cope with these potential sources of error, careful test planning must be combined with sound numerical estimation techniques.

2 TECHNICAL APPROACH

For experiments designed with a fixed dosage increment, one may obtain the mean μ and standard deviation σ by plotting the percentage of successes at each dosage level on probability paper.[5] If this data correlates well with the normal distribution, μ and σ may be extracted graphically from the plot. Unfortunately, for certain types of sensitivity tests, the dosage cannot be precisely controlled, and the data may not exactly conform to the normal distribution. The former difficulty is commonly encountered for system reliability test series. To achieve the best estimates of the mean and standard deviation for these tests, we usually apply generalized MLE procedures.[2]

2.1 Maximum Likelihood Equation

The *Method of Maximum Likelihood* is a powerful estimation procedure that was developed by R.A. Fisher during the first two decades of the Twentieth Century. To adapt this method for analyzing our sensitivity tests, we apply theoretical concepts from probability. Individual binary trials performed in the course of the test series are effectively *independent* from the standpoint of probability. The outcome (or response) of one trial has no effect on the outcome of any other trial. If we envision that the outcome of each trial is represented by its own random variable, then in terms of the dosage (an explanatory variable), these random variables are *identically distributed*. That is to say, they have the same distribution function and parameters (μ and σ). Since the trials are independent, the probability of events occurring in two trials is given by the product of their individual probabilities. Suppose that a total of n_i binary trials are conducted at the

i^{th} dosage level. Let the success probability at this level be given by p_i and the failure probability by

$$q_i = 1 - p_i . \tag{2.1.1}$$

With these assumptions, the entire test series can be envisioned as a sequence of Bernoulli (pass/fail) trials. Further suppose that there are s_i successes at this level (therefore $n_i - s_i$ failures). Then the likelihood function can be written as

$$\tilde{L} = \prod_{i=1}^{N_l} \binom{n_i}{s_i} p^{s_i} q^{n_i - s_i} , \tag{2.1.2}$$

where N_l is the number of dosage levels; the term in the parentheses is a *combination of* n_i *trials taken* s_i *at a time*.[9] The combination is used to represent the independence of order for Bernoulli trials defined at the same level. Equation (2.1.2) is the product of binomial probability distributions formulated at the dosage levels.[10] Using the natural logarithm of the likelihood function lends a great deal of convenience, i.e.,

$$L = \sum_{i=1}^{N_l} \left\{ \binom{n_i}{s_i} + (n_i - s_i)\ln(p_i) + s_i \ln(q_i) \right\}, \tag{2.1.3}$$

where $L = \ln(\tilde{L})$. By maximizing L, the natural log of the likelihood function, we effectively maximize $\tilde{L}$. Equation (2.1.3) is referred to as the log-likelihood function.

2.2 Dosage Dependant Probabilities

To estimate the "mean dosage", the dosage defined at the level where the probability equals 0.5, many probabilistic concepts must be brought together. As a

result, it is very easy to lose sight of important theoretical details. One may note that we have done nothing to connect the MLE expression to the dosage (or explanatory variable). Equation (2.1.2) is cast in the form of a binomial probability distribution, but one may note that p_i and q_i are not yet related to the dosage variable. These probabilities are related to the dosage variable though the use of a *link function*.[1] Several link functions are available, but we apply the normal probability distribution function commonly used in the "probit" method.[11] The normal distribution is based upon a continuous probability density function that must be integrated with respect to the explanatory variable (x_i at the ith level) in order to compute p_i where

$$p_i = p(x_i) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_i} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \,. \qquad (2.2.1)$$

Note that by this definition, p_i is the probability that $x \in (-\infty, x_i)$. When (2.2.1) is substituted into (2.1.3), we obtain an expression for the log-likelihood function that explicitly depends on the dosage level. To promote some simplicity, we define the variable

$$t(x) = \frac{x-\mu}{\sigma}, \qquad (2.2.2)$$

and we can rewrite p_i as

$$p_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_i} \exp\left[-\frac{t^2}{2}\right] dt \,. \qquad (2.2.3)$$

This equation is now in the form of the *standard normal distribution*.[7] Although (2.1.3), (2.2.2) and (2.2.3) are correct from the standpoint of theory, (2.2.2) must be placed in a

different form to support the estimation procedure.

2.3 Garwood's Method

In the course of this research project, the author has encountered a number of different solution procedures for (2.1.3), (2.2.2) and (2.2.3). Each procedure has its advantages and disadvantages, and in some cases, the attendant solution algorithm tends to be a little unstable. Garwood's procedure, presented below, seems to offer the most stable and robust performance for different data sets.[9] We begin this discussion with an alternate form for t, i.e.,

$$t = \frac{x-u}{\sigma} = \alpha + \beta x\,. \tag{2.3.1}$$

This expression is just a linear polynomial in x, but it offers an advantage from the standpoint of differential calculus. From (2.3.1), it is easy to show that

$$\mu = -\frac{\alpha}{\beta};\quad \sigma = \frac{1}{\beta}\,. \tag{2.3.2}$$

In the light of (2.1.3) and (2.3.1), critical points, local maxima and minima of L, may be cast in ordered pairs (α,β); they may be determined by solving the system of equations:

$$\frac{\partial L}{\partial \alpha} = 0;\quad \frac{\partial L}{\partial \beta} = 0. \tag{2.3.3}$$

By differentiating (2.1.3), we may show that

$$\frac{\partial L}{\partial \theta} = \sum_{i=1}^{N_l} \frac{n_i}{p_i q_i}\left(q_i - \frac{s_i}{n_i}\right)\frac{\partial p_i}{\partial \theta} = 0\,, \tag{2.3.4}$$

where $\theta \in (\alpha, \beta)$. Observe that

$$\frac{\partial p_i}{\partial \alpha} = \frac{\partial}{\partial \alpha}(p(t_i)) = \frac{\partial t_i}{\partial \alpha} p'(t_i), \tag{2.3.5}$$

and

$$\frac{\partial p_i}{\partial \beta} = \frac{\partial}{\partial \beta}(p(t_i)) = \frac{\partial t_i}{\partial \beta} p'(t_i). \tag{2.3.6}$$

The prime ($'$) symbol denotes differentiation with respect to the principal argument. Another operation called *differentiation with respect to a parameter* may be applied to (2.2.3) to obtain

$$p'(t_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t_i^2}{2}\right) = z_i, \tag{2.3.7}$$

where the notation z_i has been used for brevity. It is very easy to show via (2.3.1) that

$$\frac{\partial p_i}{\partial \alpha} = z_i; \quad \frac{\partial p_i}{\partial \beta} = x_i z_i. \tag{2.3.8}$$

By substituting in (2.3.4), we obtain

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{N_l} \frac{n_i}{p_i q_i}\left(q_i - \frac{s_i}{n_i}\right) z_i; \tag{2.3.9}$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{N_l} \frac{n_i}{p_i q_i}\left(q_i - \frac{s_i}{n_i}\right) x_i z_i. \tag{2.3.10}$$

Following Garwood[9], we note that $p_i = p(t_i)$ and $q_i = q(t_i)$; therefore, we can rewrite (2.3.9) and (2.3.10), respectively as

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{N_l} \zeta_i; \tag{2.3.11}$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{N_l} \zeta_i x_i , \tag{2.3.12}$$

where

$$\zeta_i = \zeta(t_i) = \frac{n_i z_i}{p_i q_i}\left(q_i - \frac{s_i}{n_i} \right). \tag{2.3.13}$$

We also need the second partial derivatives of the log-likelihood function L; hence,

$$\frac{\partial^2 L}{\partial \alpha^2} = \frac{\partial}{\partial \alpha}\sum_{i=1}^{N_l} \zeta_i = \sum_{i=1}^{N_l} \frac{\partial \zeta_i}{\partial \alpha} = \sum_{i=1}^{N_l} \frac{\partial t_i}{\partial \alpha}\frac{d\zeta_i}{dt_i} = \sum_{i=1}^{N_l} \frac{\partial t_i}{\partial \alpha}\zeta_i' . \tag{2.3.14}$$

From (2.3.1), we can easily see that

$$\frac{\partial t_i}{\partial \alpha} = 1; \quad \frac{\partial t_i}{\partial \beta} = x_i . \tag{2.3.15}$$

Hence,

$$\frac{\partial^2 L}{\partial \alpha^2} = \sum_{i=1}^{N_l} \zeta_i' . \tag{2.3.16}$$

By using (2.3.15), we can also show that

$$\frac{\partial^2 L}{\partial \alpha \partial \beta} = \frac{\partial}{\partial \alpha}\left(\frac{\partial L}{\partial \beta} \right) = \frac{\partial}{\partial \alpha}\left(\sum_{i=1}^{N_l} x_i \zeta_i \right) = \frac{\partial t_i}{\partial \alpha}\left(\sum_{i=1}^{N_l} x_i \frac{d\zeta_i}{dt_i} \right) = \sum_{i=1}^{N_l} x_i \zeta_i' , \tag{2.3.17}$$

and

$$\frac{\partial^2 L}{\partial \beta^2} = \frac{\partial}{\partial \beta}\left(\frac{\partial L}{\partial \beta} \right) = \frac{\partial}{\partial \beta}\left(\sum_{i=1}^{N_l} x_i \zeta_i \right) = \sum_{i=1}^{N_l} x_i \frac{\partial t_i}{\partial \beta}\frac{d\zeta_i}{dt_i} = \sum_{i=1}^{N_l} x_i^2 \zeta_i' . \tag{2.3.18}$$

In order to evaluate (2.3.16) through (2.3.18), we must derive an expression for ζ_i'. We begin with (2.3.13).

$$\zeta_i' = \frac{d\zeta_i}{dt_i} = \frac{d}{dt_i}\left[\frac{n_i z_i}{p_i q_i}\left(q_i - \frac{s_i}{n_i} \right) \right]; \tag{2.3.19}$$

by applying the quotient rule,

$$\zeta_i' = \frac{p_i q_i \left[n_i z_i \left(q_i - s_i / n_i \right) \right]' - n_i z_i \left(q_i - s_i / n_i \right) \left(p_i q_i \right)'}{\left(p_i q_i \right)^2}; \tag{2.3.20}$$

$$\zeta_i' = \frac{n_i}{p_i q_i} \left[z_i \left(q_i - \frac{s_i}{n_i} \right)' + z_i' \left(q_i - \frac{s_i}{n_i} \right) \right] - \frac{n_i z_i}{\left(p_i q_i \right)^2} \left(q_i - \frac{s_i}{n_i} \right) \left(p_i q_i' + p_i' q_i \right). \tag{2.3.21}$$

By using (2.1.1) and (2.3.7), we can show that

$$\zeta_i' = \frac{n_i}{p_i q_i} \left[- z_i^2 + z_i' \left(q_i - \frac{s_i}{n_i} \right) \right] - \frac{n_i z_i}{\left(p_i q_i \right)^2} \left(q_i - \frac{s_i}{n_i} \right) \left(- z_i p_i + z_i q_i \right). \tag{2.3.22}$$

Further algebraic simplification yields

$$\zeta_i' = - \frac{n_i z_i^2}{p_i q_i} + \frac{n_i z_i'}{p_i q_i} \left(q_i - \frac{s_i}{n_i} \right) - \frac{n_i z_i^2}{p_i^2 q_i} \left(q_i - \frac{s_i}{n_i} \right) + \frac{n_i z_i^2}{p_i q_i^2} \left(q_i - \frac{s_i}{n_i} \right). \tag{2.3.23}$$

Hence,

$$\zeta_i' = - \frac{n_i z_i^2}{p_i q_i} + \frac{n_i z_i}{p_i q_i} \left(q_i - \frac{s_i}{n_i} \right) \left[\frac{z_i'}{z_i} - \frac{z_i}{p_i} + \frac{z_i}{q_i} \right]. \tag{2.3.24}$$

It is evident from this sequence of equations that we must evaluate z_i'. We may do so easily by differentiating (2.3.7), i.e.,

$$z_i' = - \frac{t_i}{\sqrt{2\pi}} \exp\left(- \frac{t_i^2}{2} \right) = t_i z_i. \tag{2.3.25}$$

By substituting (2.3.24) and (2.3.25) into equations (2.3.16) through (2.3.18), we obtain Garwood's formulas for $\partial^2 L / \partial \alpha^2$, $\partial^2 L / \partial \alpha \partial \beta$ and $\partial^2 L / \partial^2 \beta$.

2.4 An Alternative Method

Garwood's method provides a set of exact formulas for the partial derivatives of the log-likelihood function with respect to parameters α and β.[9] We have successfully employed these formulas to solve test problems, particularly those regarding Garwood's examples. Unfortunately, problems arise when this method is applied to sensitivity tests that are comprised of a mixture of binomial and binary trials. For endpoint probabilities, those near zero and unity, terms such as ζ_i and ζ_i' become undefined due the presence of the factor $p_i q_i$ existing in the denominator. For endpoint probabilities, $p_i q_i = 0$; this situation is routinely encountered for levels in the explanatory variable characterized by a single binary trial (or its binomial equivalent). After observing this difficulty, we decided to revisit the derivation and search for an alternative form of derivatives $\partial L/\partial\alpha$, $\partial L/\partial\beta$, $\partial^2 L/\partial\alpha^2$, $\partial^2 L/\partial\alpha\partial\beta$ and $\partial^2 L/\partial\beta^2$. An ideal alternative formulation is easier to control near the endpoints.

To begin our derivation, we differentiate equation (2.1.3) as follows.

$$\frac{\partial L}{\partial \alpha} = \frac{\partial}{\partial \alpha}\sum_{i=1}^{N_l}\left[\binom{n_i}{s_i} + (n_i - s_i)\ln p_i + s_i \ln q_i\right]. \tag{2.4.1}$$

If we apply the derivative term by term through the summation, we obtain

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{N_l}\left[(n_i - s_i)\frac{\partial \ln p_i}{\partial \alpha} + s_i \frac{\partial \ln q_i}{\partial \alpha}\right]; \tag{2.4.2}$$

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{N_l}\left[(n_i - s_i)\frac{\partial t_i}{\partial \alpha}\frac{d \ln p_i}{dt_i} + s_i \frac{\partial t_i}{\partial \alpha}\frac{d \ln q_i}{dt_i}\right]. \tag{2.4.3}$$

By taking derivatives of the natural logarithm, we have that

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{N_I} \frac{\partial t_i}{\partial \alpha}\left[\left(n_i - s_i\right)\frac{1}{p_i}\frac{d\,p_i}{dti} + s_i \frac{1}{q_i}\frac{d\,q_i}{dt} \right], \tag{2.4.4}$$

and this expression can be simplified by using (2.1.1), (2.3.15) and (2.3.25), i.e.,

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{N_I}\left[\left(n_i - s_i\right)\frac{z_i}{p_i} - s_i \frac{z_i}{q_i} \right]. \tag{2.4.5}$$

By similar means, we can show that

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{N_I} x_i \left[\left(n_i - s_i\right)\frac{z_i}{p_i} - s_i \frac{z_i}{q_i} \right]. \tag{2.4.6}$$

The form of equations (2.4.5) and (2.4.6) is quite interesting; the summand contains two terms, and each of these terms contains a ratio of the probability density function to its cumulative distribution function both evaluated at t_i (or x_i under inverse transformation). The behavior of the first (second) term $\sim z_i / p_i$ is questionable as $p_i \to 0$. Conversely, the behavior of the second term ($\sim z_i / q_i$) requires investigation as $p_i \to 1$ (or $q_i \to 0$). If we can assert that z_i / p_i tends to zero as $p_i \to 0$, we can exclude this term from (2.4.5) in the limit. Consider the following claim.

Claim P-1: *Let* z be *the normal probability density function, and let* p *be the cumulative distribution function both defined in independent variable* t. *The ratio* z / p *is well defined and tends to zero as* $p \to 0$.

Justification: By examining equation (2.3.7) and (2.2.3), respectively, we see that both z *and* p *approach zero if and only if* $t \to -\infty$. *The ratio* z / p *may be written as follows.*

$$\frac{z}{p}(t) = \frac{\exp\left(-\frac{t^2}{2}\right)}{\int_{-\infty}^{t} \exp\left(-\frac{\tilde{t}^2}{2}\right) d\tilde{t}}. \qquad \text{(P-1.1)}$$

Clearly, the Taylor series expansion for the exponential function,

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots, \qquad \text{(P-1.2)}$$

is a convergent power series for $x < 0$. Hence, it flows that

$$\exp\left(-\frac{t^2}{2}\right) = 1 - \frac{t^2}{2 \cdot 1!} + \frac{t^4}{4 \cdot 2!} - \frac{t^6}{8 \cdot 3!} + \cdots = \sum_{i=0}^{n} S_i^1 + O\left(t^{2n}\right) \qquad \text{(P-1.3)}$$

is also a convergent power series for any real valued t. To evaluate p, as is specified in (P-1.1) we compute the anti-derivative of (P-1.3) term by term, i.e.,

$$\int \exp\left(-\frac{\tilde{t}^2}{2}\right) d\tilde{t} = t - \frac{t^3}{6 \cdot 1!} + \frac{t^5}{20 \cdot 2!} - \frac{t^7}{56 \cdot 3!} + \cdots; \qquad \text{(P-1.4)}$$

$$\int \exp\left(-\frac{\tilde{t}^2}{2}\right) d\tilde{t} = t\left[1 - \frac{t^2}{6 \cdot 1!} + \frac{t^4}{20 \cdot 2!} - \frac{t^6}{56 \cdot 3!} + \cdots\right] = t \sum_{n} S_i^2 + O\left(t^{2n}\right). \qquad \text{(P-1.5)}$$

$\sum_n S_n$ is the alternating power series contained with the brackets of (P-1.5). If we apply the ratio test to this series, we obtain

$$\left|\frac{S_{n+1}}{S_n}\right| = \frac{t^{2(n+1)}}{(n+1)!(2(n+1)+1)(2(n+2))} \cdot \frac{n!(2n+1)(2(n+1))}{t^{2n}}. \qquad \text{(P-1.6)}$$

Algebraic simplification yields that

$$\left|\frac{S_{n+1}}{S_n}\right| = t^2 \cdot \frac{2n+1}{(2n+3)(n+2)} = t^2 \cdot \frac{2n+1}{2n^2+7n+5}, \qquad \text{(P-1.7)}$$

and thus,

$$\lim_{n\to\infty}\left|\frac{S_{n+1}}{S_n}\right| = 0\,. \tag{P-1.8}$$

Therefore, by the ratio test, the power series in (P-1.5) is convergent, therefore finite, for any real value of t. It is also non-zero for finite t. Since both of the series in (P-1.1) are convergent, they both have finite limits, say $f(t)$ for the numerator and $g(t)$ for the denominator, so for any particular real t

$$\frac{z}{p} = \frac{f(t)}{\tilde{t}\,g(\tilde{t})\Big|_{-\infty}^{t}} = \lim_{\tilde{t}\to-\infty}\frac{f(t)}{t\,g(t)-\tilde{t}\,g(\tilde{t})} = \frac{f(t)}{t\,g(t)} \tag{P-1.9}$$

given the properties of the cumulative distribution function as $t\to-\infty$. Since $f(t)$ and $g(t)$ are the finite limits of series that converge with the same order, we have that

$$\lim_{p\to 0}\frac{z}{p} = \lim_{t\to-\infty}\frac{f(t)}{t\,g(t)} = 0\,. \ \square \tag{P-1.10}$$

By using a proof similar to that shown in Proposition P-1, the ratio z/q also approaches zero as $p\to 1$ ($q\to 0$). It follows that (2.4.5) and (2.4.6) can be expressed

as

$$\frac{\partial L}{\partial \alpha} = \begin{cases} -\sum_{i=1}^{N_l} s_i\, z_i\,, & p_i = 0 \\ \sum_{i=1}^{N_l}(n_i - s_i)z_i, & q_i = 0 \end{cases} \tag{2.4.7}$$

and

$$\frac{\partial L}{\partial \beta} = \begin{cases} -\sum_{i=1}^{N_l} x_i s_i z_i , & p_i = 0 \\ \sum_{i=1}^{N_l} x_i (n_i - s_i) z_i , & q_i = 0 \end{cases} \tag{2.4.8}$$

By using these equations for endpoint probabilities, $\partial L / \partial \alpha$ and $\partial L / \partial \beta$, (elements of the Jacobian matrix for the log-likelihood function) remain well defined for all possible values of p_i and q_i.

In addition to the Jacobian matrix, we must also use the matrix of second partial derivatives or Hessian matrix. The reason for its necessity will be made clear in the next section. We would like to construct the second partial derivatives in terms of the ratios z_i / p_i and z_i / q_i. To derive $\partial^2 L / \partial \alpha^2$, we differentiate (2.4.5).

$$\frac{\partial^2 L}{\partial \alpha^2} = \sum_{i=1}^{N_l} \left[(n_i - s_i) \frac{\partial}{\partial \alpha} \left(\frac{z_i}{p_i} \right) - s_i \frac{\partial}{\partial \alpha} \left(\frac{z_i}{q_i} \right) \right]. \tag{2.4.9}$$

With some careful mathematics and the use of (2.2.3), (2.3.7) and (2.3.15), we can show that

$$\frac{\partial}{\partial \alpha} \left(\frac{z_i}{p_i} \right) = \frac{\partial t_i}{\partial \alpha} \frac{d}{dt_i} \left(\frac{z_i}{p_i} \right) = -t_i \left(\frac{z_i}{p_i} \right) - \left(\frac{z_i}{p_i} \right)^2 ; \tag{2.4.10}$$

$$\frac{\partial}{\partial \alpha} \left(\frac{z_i}{q_i} \right) = \frac{\partial t_i}{\partial \alpha} \frac{d}{dt_i} \left(\frac{z_i}{q_i} \right) = -t_i \left(\frac{z_i}{q_i} \right) + \left(\frac{z_i}{q_i} \right)^2 . \tag{2.4.11}$$

After substituting (2.4.10) and (2.4.11) into (2.4.9), we obtain

$$\frac{\partial^2 L}{\partial \alpha^2} = \sum_{i=1}^{N_l} \left[(n_i - s_i) \left(\frac{z_i}{p_i} \right) \left(-t_i - \frac{z_i}{p_i} \right) - s_i \left(\frac{z_i}{q_i} \right) \left(-t_i + \frac{z_i}{q_i} \right) \right]. \tag{2.4.12}$$

Note that (2.4.12) is characterized by the presence of the ratios z_i / p_i and z_i / q_i but no other terms containing z_i or p_i. Now let us consider the mixed partial derivative

$$\frac{\partial^2 L}{\partial\alpha\partial\beta} = \sum_{i=1}^{N_l}\left[(n_i - s_i)\frac{\partial}{\partial\beta}\left(\frac{z_i}{p_i}\right) - s_i\frac{\partial}{\partial\beta}\left(\frac{z_i}{q_i}\right)\right]. \tag{2.4.13}$$

By carefully differentiating and using previously derived results, we can show that

$$\frac{\partial}{\partial\beta}\left(\frac{z_i}{p_i}\right) = \frac{\partial t_i}{\partial\beta}\frac{d}{dt_i}\left(\frac{z_i}{p_i}\right) = -x_i\left(t_i\left(\frac{z_i}{p_i}\right) + \left(\frac{z_i}{p_i}\right)^2\right), \tag{2.4.14}$$

and

$$\frac{\partial}{\partial\beta}\left(\frac{z_i}{q_i}\right) = \frac{\partial t_i}{\partial\beta}\frac{d}{dt_i}\left(\frac{z_i}{q_i}\right) = -x_i\left(t_i\left(\frac{z_i}{q_i}\right) - \left(\frac{z_i}{q_i}\right)^2\right). \tag{2.4.15}$$

By substituting (2.4.14) and (2.4.15) into (2.4.13), we can show that

$$\frac{\partial^2 L}{\partial\alpha\partial\beta} = -\sum_{i=1}^{N_l} x_i\left[(n_i - s_i)\left(\frac{z_i}{p_i}\right)\left(t_i + \frac{z_i}{p_i}\right) - s_i\left(\frac{z_i}{q_i}\right)\left(t_i - \frac{z_i}{q_i}\right)\right], \tag{2.4.16}$$

and again the desired ratios have been preserved. Recalling (2.4.6), the remaining partial derivative is given by

$$\frac{\partial^2 L}{\partial\beta^2} = \frac{\partial}{\partial\beta}\left[\sum_{i=1}^{N_l} x_i\left\{(n_i - s_i)\frac{z_i}{p_i} - s_i\frac{z_i}{q_i}\right\}\right]; \tag{2.4.17}$$

hence,

$$\frac{\partial^2 L}{\partial\beta^2} = \sum_{i=1}^{N_l} x_i\left[(n_i - s_i)\frac{\partial}{\partial\beta}\left(\frac{z_i}{p_i}\right) - s_i\frac{\partial}{\partial\beta}\left(\frac{z_i}{q_i}\right)\right]. \tag{2.4.18}$$

After substituting (2.4.14) and (2.4.15), we have that

$$\frac{\partial^2 L}{\partial \beta^2} = \sum_{i=1}^{N_l} x_i \left[(n_i - s_i) \left\{ -x_i \left(t_i \left(\frac{z_i}{p_i} \right) + \left(\frac{z_i}{p_i} \right)^2 \right) \right\} - s_i \left\{ -x_i \left(t_i \left(\frac{z_i}{q_i} \right) - \left(\frac{z_i}{q_i} \right)^2 \right) \right\} \right]; \quad (2.4.19)$$

or after simplifying,

$$\frac{\partial^2 L}{\partial \beta^2} = -\sum_{i=1}^{N_l} x_i^2 \left[(n_i - s_i) \left(\frac{z_i}{p_i} \right) \left(t_i + \frac{z_i}{p_i} \right) - s_i \left(\frac{z_i}{q_i} \right) \left(t_i - \frac{z_i}{q_i} \right) \right]. \quad (2.4.20)$$

To recap, this section has presented an alternative formulation of Garwood's method that is useful for evaluating endpoint probabilities. The elements of the Jacobian matrix are given by (2.4.5) and (2.4.6) for intermediate (non-endpoint, i.e., $0 < p < 1$) probabilities while the endpoint formulas are given by (2.4.7) and (2.4.8). Equations (2.4.12), (2.4.16) and (2.4.20) contain the elements for the Hessian matrix. These formulas are immediately suitable for calculating intermediate probabilities. To calculate endpoint probabilities, the same formulas are easily modified in the manner used to derive (2.4.7) and (2.4.8).

2.5 Newton's Solution Method

In Sections 2.3 and 2.4, it was demonstrated that we may fit sensitivity test data to the normal distribution through Fisher's Method of Maximum Likelihood.[12] Moreover, the procedure requires that we estimate two parameters, the distribution mean and standard deviation, μ and σ, respectively. To do so, we must determine α and β satisfying (2.3.3). Estimates of α and β are obtained by solving equations (2.3.3) cast either in Garwood's form (Section 2.3) or an alternative form (Section 2.4). The fitting procedure takes the form of an iterative scheme; rewrite α and β as follows.

$$\alpha = \alpha_0 + \Delta\alpha\,; \quad \beta = \beta_0 + \Delta\beta\,. \tag{2.5.1}$$

Substitute (2.5.1) into (2.3.3) and apply Taylor's series for two variables; observe that

$$\frac{\partial L}{\partial \alpha}(\alpha,\beta) = \frac{\partial L}{\partial \alpha}(\alpha_0,\beta_o) + \Delta\alpha\,\frac{\partial^2 L}{\partial \alpha^2}(\alpha_0,\beta_0) + \Delta\beta\,\frac{\partial^2 L}{\partial \alpha \partial \beta}(\alpha_0,\beta_0) + O\left(\Delta^2\right); \tag{2.5.2}$$

$$\frac{\partial L}{\partial \beta}(\alpha,\beta) = \frac{\partial L}{\partial \beta}(\alpha_0,\beta_0) + \Delta\alpha\,\frac{\partial^2 L}{\partial \alpha \partial \beta}(\alpha_0,\beta_0) + \Delta\beta\,\frac{\partial^2 L}{\partial \beta^2}(\alpha_0,\beta_0) + O\left(\Delta^2\right), \tag{2.5.3}$$

where the truncation error has order

$$\Delta^2 = \phi(\Delta a)^2 \;+\; \theta(\Delta\alpha)(\Delta\beta) \;+\; \psi(\Delta\beta)^2\,,$$

for real numbers ϕ, θ and ψ. If we assume that α and β satisfy (2.3.3) within $O\left(\Delta^2\right)$, we can rewrite (2.5.2) and (2.5.3) in matrix form, i.e.,

$$\begin{bmatrix} \dfrac{\partial^2 L}{\partial \alpha^2}(\alpha_0,\beta_0) & \dfrac{\partial^2 L}{\partial \alpha \partial \beta}(\alpha_0,\beta_0) \\ \dfrac{\partial^2 L}{\partial \alpha \partial \beta}(\alpha_0,\beta_0) & \dfrac{\partial^2 L}{\partial \beta^2}(\alpha_0,\beta_0) \end{bmatrix} \begin{bmatrix} \Delta\alpha \\ \\ \Delta\beta \end{bmatrix} = -\begin{bmatrix} \dfrac{\partial L}{\partial \alpha}(\alpha_0,\beta_0) \\ \\ \dfrac{\partial L}{\partial \beta}(\alpha_0,\beta_0) \end{bmatrix}. \tag{2.5.4}$$

The 2x2 matrix on the left side of (2.5.4) is H, the Hessian matrix, and the Jacobian matrix (a vector in this case) resides on the right side. If we envision (α_0,β_0) as being a starting "guess", then $(\Delta\alpha,\Delta\beta)$ is an increment that when added to (α_0,β_0) tends to improve the accuracy of the estimate. We can solve (2.5.4) for increments by inverting H and multiplying from the left. The inverse matrix H^{-1} is easily obtained for a 2x2 matrix insofar as $\left|H(\alpha_0,\beta_0)\right| \neq 0$. Let $\vec{J}$ represent the Jacobian vector in (2.5.4); then $(\Delta\alpha,\Delta\beta)$ is given by

$$[\Delta\alpha,\Delta\beta]^T = H^{-1}(\alpha_0,\beta_0)\cdot\vec{J}(\alpha_0,\beta_0). \tag{2.5.5}$$

From (2.5.1), it follows that

$$[\alpha,\beta]^T = [\alpha_0,\beta_0]^T + H^{-1}(\alpha_0,\beta_0)\cdot \vec{J}(\alpha_0,\beta_0)\ , \qquad (2.5.6)$$

where $[\alpha,\beta]^T$ contains the refined parameter estimates. As with other Newton approximation methods, (2.5.6) is easily structured as an iterative scheme. After a number of iterations, when (2.5.6) has been converged to the desired level of accuracy, the distribution mean and standard deviation can be calculated from (2.3.2).

2.6 Confidence Intervals

The confidence interval represents one of the more arcane concepts in statistics. In many scientific problems, we are interested in estimating some unknown parameter, say the "ideal" dosage for an antibiotic. When the ideal dose is administered, it should establish a concentration in tissue that is suitable for the eradication of bacteria. Given the amount of variation that is routinely created in the manufacturing process, we would like to estimate an effective range for the dosage and attach a probabilistic level of assurance to it. This dosage range is known as a *confidence interval*, and the process of its determination is known as an *interval estimate.* The *end points* of the interval act as *random variables* while the statistical parameter contained within the interval is regarded as *fixed, but unknown*.[12] Now let us illustrate where the situation becomes confusing.

Consider a physical or biological process that exhibits some randomness or "noise". Suppose that the behavior of the process may be characterized, in part, by an unknown, yet constant *non-random* parameter, say θ. (The mean of a statistical

distribution is a good example of this type of parameter). Unfortunately in most cases, θ cannot be measured with exactness. Instead, it must be estimated by conducting a series of independent experimental trials. We denote θ's estimate as $\hat{\theta}$. Based upon the information provided by the experiments, we can also obtain an interval estimate $[\theta_L, \theta_H]$ associated with $\hat{\theta}$. At this point the investigator would like to obtain a probability p for the interval and then state,

"With $100p$ *% certainty, the real value of* θ *is contained within* $[\theta_L, \theta_H]$*."*

In truth, it is *not possible* to make this claim.[10] Why? As it happens, the interval endpoints θ_L and θ_H are calculated based upon the value $\hat{\theta}$. At the conclusion of the experimental measurements, $\hat{\theta}$ is known, and θ_L and θ_H can be calculated directly. As they pertain to a single series of experiments, $\hat{\theta}$, θ_L and θ_H are fixed; they have *no random behavior*. Hence, we cannot assign a probability to any of these values, so the claim made above cannot be true. It is important that we understand what probability really means for this situation.

In the mathematics of probability, the interval endpoints are represented by random variables. Let these random variables be denoted as Θ_L and Θ_R. If we repeat the *entire* experimental series a number of times, we would obtain a series of values for these random variables. For example, the first test series may produce an estimate $\hat{\theta}$ and the values $\Theta_L = \theta_L$ and $\Theta_R = \theta_R$ as discussed earlier. The probabilistic aspects of the interval only become evident when we think of repeating the test series a large number of times, obtaining a series of values for Θ_L and Θ_R. In this context, both Θ_L

and Θ_R exhibit random fluctuations, so it makes sense to assign probabilities to their outcomes.[12] For the first test series, let us suppose that we calculate an interval $[\theta_L, \theta_R]$ with a probability of 95%. This result is interpreted as follows. Imagine that we repeat the entire test series an additional 99 times and compute a different $[\theta_L, \theta_R]$. Then 95 out of these 100 intervals will contain the real parameter θ. We can never calculate the probability that θ lies inside of a *single* interval $[\theta_L, \theta_R]$. Now let us discuss different types of confidence intervals. Although our treatment of confidence intervals is not rigorous, we hope to convey some of the mathematics behind the construction of confidence intervals. In most cases, we will draw upon sampling theory as our core resource.

2.6.1 The Success Ratio Confidence Interval

For the success ratio (or success probability) confidence interval, we actually estimate a success ratio $\hat{p}$ and calculate a confidence interval $[p_L, p_R]$ around this probability. This interval is determined with a confidence level of $100(1-\alpha)\%$, $\alpha < 1$. Hence, for $100(1-\alpha)$ times out of 100, we expect that this interval will contain the real success ratio p. We can derive the endpoint probabilities for this interval by using theory associated with the binomial distribution.[13] Begin by invoking the DeMoivre-Laplace Theorem, i.e., for large values of n (the number of trials per experimental series), the ratio

$$z = \frac{\frac{Y}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \qquad (2.6.1.1)$$

tends towards the standard normal distribution $N(0,1)$. The similarity of (2.6.1.1) to $N(0,1)$ can be illustrated by realizing that for the binomial distribution, $\mu = np$ and $\sigma^2 = np(1-p)$.[13] With these substitutions, (2.6.1.1) can be rewritten, i.e.,

$$z = \frac{Y - \mu}{\sigma}. \qquad (2.6.1.2)$$

This expression is identical to (2.2.2), the argument of the standard normal distribution.

$N(0,1)$ is a symmetric function about its zero mean, so a typical confidence interval $(-z_{\alpha/2}, z_{\alpha/2})$ is also symmetric about the mean. The probability associated with this interval is obtained by integrating the normal density function over $\left(-z_{\alpha/2}, z_{\alpha/2}\right)$. The excluded regions $(-\infty, -z_{\alpha/2})$ and $(z_{\alpha/2}, \infty)$ form the "tails" of this distribution. The "tails" are important when determining the probability or *confidence coefficient* associated with the interval. Let the combined probability (area under the curve) for the "tails" equal α; then the confidence coefficient for the interval $\left(-z_{\alpha/2}, z_{\alpha/2}\right)$ is given by $100(1-\alpha)\%$. As an example, consider a 95% confidence coefficient; the attendant value of α is 0.05. It follows that the area under each section of the "tail" is $\alpha/2 = 0.025$. The interval endpoints on the explanatory (z) axis for $N(0,1)$ can now be identified by solving for $-z_{\alpha/2}$ in the equation

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-z_{\alpha/2}} \exp\left(-\frac{z^2}{2}\right) dz = \frac{\alpha}{2}. \qquad (2.6.1.3)$$

It follows that the confidence level for the success ratio interval is given by[13]

$$\text{Prob}\left[-z_{\alpha/2} < \frac{\frac{Y}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} < z_{\alpha/2} \right] \cong 1-\alpha \,. \qquad (2.6.1.4)$$

Let us carefully consider the meaning of (2.6.1.4). The *true* success ratio is given by the unknown p while its *estimate* is given by the measured value Y/n (later referred to as $\hat{p}$). If we can "solve for p" within the restrictions of (2.6.1.4), we automatically obtain an interval estimate for p with the confidence coefficient $(1-\alpha)$. Following Larsen and Marx[13], we can rewrite the argument of (2.6.1.4) as

$$\frac{\left|\frac{Y}{n} - p\right|}{\sqrt{\frac{p(1-p)}{n}}} < z_{\alpha/2} \,. \qquad (2.6.1.5)$$

By squaring and rearranging terms, we obtain

$$\left(\frac{Y}{n} - p\right)^2 - z_{\alpha/2}^2 \left[\frac{p(1-p)}{n}\right] < 0 \,. \qquad (2.6.1.6)$$

Albeit with the use of tedious algebraic manipulations, we can rewrite (2.6.1.6) as

$$p^2\left(1+\frac{z_{\alpha/2}^2}{n}\right) - p\left(2\frac{Y}{n} + \frac{z_{\alpha/2}^2}{n}\right) + \left(\frac{Y}{n}\right)^2 < 0 \,. \qquad (2.6.1.7)$$

Equation (2.6.1.7), when viewed as an equality, represents a parabola.[13] The roots of this equation serve to delimit regions where the locus of the parabola is greater or less than zero. In that sense, the roots of (2.6.1.7) are the endpoints of the confidence interval. These endpoints, denoted p_{Low} and p_{High}, for the success probability's

$100(1-\alpha)\%$ confidence interval may be determined through the use of the quadratic formula, i.e.,

$$p_{Low} = \frac{\frac{Y}{n} + \frac{z_{\alpha/2}^2}{2n} - \frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\frac{Y}{n}\left(1-\frac{Y}{n}\right)+\frac{z_{\alpha/2}^2}{4n}}}{1+\frac{z_{\alpha/2}^2}{n}}; \qquad (2.6.1.8)$$

$$p_{High} = \frac{\frac{Y}{n} + \frac{z_{\alpha/2}^2}{2n} + \frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\frac{Y}{n}\left(1-\frac{Y}{n}\right)+\frac{z_{\alpha/2}^2}{4n}}}{1+\frac{z_{\alpha/2}^2}{n}}. \qquad (2.6.1.9)$$

These root formulas deserve further comment. You may recall that the confidence coefficient was based upon $N(0,1)$ with the symmetric interval $(-z_{\alpha/2}, z_{\alpha/2})$. By rebuilding the Laplace-DeMoivre relationship (2.6.1.1), we have developed a success ratio interval that is symmetric about the quantity

$$\tilde{p} = \frac{\frac{Y}{n} + \frac{z_{\alpha/2}^2}{2n}}{1+\frac{z_{\alpha/2}^2}{n}}, \qquad (2.6.1.10)$$

where Y/n can be thought of as a measured success ratio or success probability (however crude). It is interesting to note that Y/n does not lie at the center of the interval. Instead, it has been translated and scaled. When n is large, $z_{\alpha/2}^2/n$ is small, so $\tilde{p} \cong Y/n$; in these cases, the Laplace-DeMoivre theorem strictly applies.

2.6.2 The Success Ratio Confidence Interval for an Explanatory Variable

For sensitivity testing, the binomial distribution remains a critical mathematical construct for determining the statistical distribution of success for a random system (physical, biological or otherwise). However, the binomial distribution alone lacks the ability to connect the system's response to an explanatory variable. This deficiency is remedied by replacing the success probability p with a probability calculated from the cumulative distribution function. This function is defined in terms of an explanatory variable. See Section 2 of this report for the details. By doing so, the probability of success can be calculated with respect to changes in the explanatory variable. As a result, our confidence intervals now lie on the explanatory variable axis instead of on the probability locus $[0,1]$. To illustrate this concept, let us reconsider an example that we began discussing in Section 2.6.

Our example addresses a problem in drug efficacy formulated in terms of a question. At what "ideal" dosage will an antibiotic achieve curative blood concentration with a chosen level or reliability? This dosage (or a suitable mathematical transformation of it) serves as the explanatory variable. A series of sensitivity tests that vary dosage are conducted in order to calibrate a normal distribution function for the drug's success. In common practice, we attempt to estimate the dosage at which a drug will successfully cure 99% of the test subjects. We are also interested in computing a confidence interval for dosage with a confidence coefficient of 95%. This confidence interval is interpreted as follows. If we were to conduct the entire series of sensitivity tests 100 times and determine a dosage confidence interval for each, the *actual dosage*

that cures 99% of the test subjects in contained in 95 out of 100 of the intervals. As it happens, we can estimate this interval based upon the theory presented in the preceding section.

If we revisit equations (2.6.1.8) and (2.6.1.9), we notice the presence of the term Y/n. In truth, this term is the number of successful trials divided by the total number of trials for the test series. This fraction is a crude success probability; it is analogous to the probability of antibiotic efficacy ($\hat{p}=0.99$) that we would like to see. If we substitute $\hat{p}$ for Y/n, we obtain

$$p_{Low}=\frac{\hat{p}+\frac{z_{\alpha/2}^2}{2n}-\frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})+\frac{z_{\alpha/2}^2}{4n}}}{1+\frac{z_{\alpha/2}^2}{n}}; \qquad (2.6.2.1)$$

$$p_{High}=\frac{\hat{p}+\frac{z_{\alpha/2}^2}{2n}+\frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})+\frac{z_{\alpha/2}^2}{4n}}}{1+\frac{z_{\alpha/2}^2}{n}}. \qquad (2.6.2.2)$$

By again noting that the mean and variance for the binomial distribution are given by

$$\mu=np; \qquad (2.6.2.3)$$

$$\sigma^2=np(1-p), \qquad (2.6.2.4)$$

we can substitute (2.6.2.4) into p_{Low} and p_{High} to obtain

$$p_{Low}=\frac{\hat{p}+\frac{z_{\alpha/2}^2}{2n}-\frac{z_{\alpha/2}}{n}\sqrt{\sigma^2+\frac{z_{\alpha/2}^2}{4}}}{1+\frac{z_{\alpha/2}^2}{n}}; \qquad (2.6.2.5)$$

$$p_{High} = \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} + \frac{z_{\alpha/2}}{n}\sqrt{\sigma^2 + \frac{z_{\alpha/2}^2}{4}}}{1 + \frac{z_{\alpha/2}^2}{n}}. \qquad (2.6.2.6)$$

These equations cast a probability interval around $\hat{p}$, so the only task remaining is to map p_{Low} and p_{High} onto corresponding points on the explanatory variable axis. From (2.2.1), we may obtain the dosage interval's endpoints by solving the following equations for ρ_{Low} and ρ_{High}.

$$p_{Low} = \frac{1}{\hat{\sigma}\sqrt{2\pi}} \int_{-\infty}^{\rho_{Low}} \exp\left[-\frac{1}{2}\left(\frac{\rho - \hat{\mu}}{\hat{\sigma}} \right)^2 \right] d\rho ; \qquad (2.6.2.7)$$

$$p_{High} = \frac{1}{\hat{\sigma}\sqrt{2\pi}} \int_{-\infty}^{\rho_{High}} \exp\left[-\frac{1}{2}\left(\frac{\rho - \hat{\mu}}{\hat{\sigma}} \right)^2 \right] d\rho , \qquad (2.6.2.8)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the maximum likelihood estimates for the mean and standard deviation obtained by using the techniques described in Section 2.4.

2.6.3 Confidence Interval for the Mean

The techniques documented within this report are designed to determine the statistical distribution associated with the results of a series of sensitivity tests. For the drug efficacy problem, we have assumed *a priori* that the data fits a normal distribution $N(\mu,\sigma)$. The purpose of our analysis is to estimate the values of μ and σ, the two parameters that determine the distribution. It is worthwhile to restate that μ and σ are parameters, *not random variables*. Hence, they have fixed, but unknown values. Imagine that we estimate each of these parameters by random sampling; then the

sample mean and sample variance are represented by random variables associated with random intervals. It follows that interval estimation is based upon statistics associated with *sampling* the distribution. We must imagine that we are estimating the mean of the normal distribution through sampling in order to calculate its confidence interval.

For n samples x_i taken from a $N(\mu,\sigma^2)$ distribution, the sample mean and variance are respectively defined as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \; ; \tag{2.6.3.1}$$

$$S^2 \; = \; \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^2 \; . \tag{2.6.3.2}$$

The sampling mean is known to have the distribution $N(\mu,\sigma^2/n)$.[12] As a result, the random variable

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{2.6.3.3}$$

has the standard normal distribution $N(0,1)$. For estimating the mean's confidence interval, we assume that the standard deviation σ is known.[12] Since this distribution is

symmetric around a zero mean, consider the interval $(-a,a)$ on the axis, i.e.,

$$-a \; \le \; \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \; \le \; a \, . \tag{2.6.3.4}$$

We can obtain a similar relationship for μ if we first multiply (2.6.3.3) by -1, i.e.,

$$a \geq \frac{\mu - \overline{X}}{\sigma/\sqrt{n}} \geq -a. \qquad (2.6.3.5)$$

By algebraically manipulating (2.6.3.5), we can show that

$$\overline{X} + \frac{\sigma a}{\sqrt{n}} \geq \mu \geq \overline{X} - \frac{\sigma a}{\sqrt{n}}. \qquad (2.6.3.6)$$

(2.6.3.6) is a random interval (since $\hat{X}$ is a random variable) associated with the real mean for X. From the equivalence of (2.6.3.4) and (2.6.3.6), we can conclude that

$$\text{Prob}\left[\overline{X} + \frac{\sigma a}{\sqrt{n}} \geq \mu \geq \overline{X} - \frac{\sigma a}{\sqrt{n}}\right] = \text{Prob}\left[a \geq \frac{\mu - \overline{X}}{\sigma/\sqrt{n}} \geq -a\right]. \qquad (2.6.3.7)$$

The probability on the right side of (2.6.3.7) is given by the area under the standard normal curve. Denote this area by $1-\alpha$. In order to determine the limits for the confidence interval, we must derive a relationship between a and α.

Since the area under the standard normal curve is $1-\alpha$, the area under the "tails" of the distribution is equal α. The "tails" are symmetric, so each "tail" is associated with the probability $\alpha/2$. Along the abscissa of the standard normal distribution, the "tails" lie in the intervals $(-\infty, -a)$ and (a, ∞). It follows that the value of a is determined from solving the equation

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-a} \exp\left(-\frac{z^2}{2}\right) dz = \frac{\alpha}{2}. \qquad (2.6.3.8)$$

When a has been computed, it may be combined with a single estimate $\bar{x}$ of the sample mean to compute the confidence interval

$$\left(\bar{x} - \frac{a\sigma}{\sqrt{n}},\ \bar{x} + \frac{a\sigma}{\sqrt{n}} \right) \qquad (2.6.3.9)$$

with confidence coefficient $1-\alpha$ and a given by the solution of (2.6.3.8).

Equation (2.6.3.9) presents intriguing connotations when examined in the context of sensitivity testing. The methods in Sections 2.4 and 2.5 fit go/no-go data to a normal distribution while (2.3.3.8) and (2.6.3.9) are designed for sampling a normal distribution. The distinction is illusory and subtle, but we must examine it in the interest of full disclosure. Let us modify John A. Wheeler's famous dictum and create a statistically testable hypothesis, i.e.,[14]

Aging male computational physicists have no hair.

If we were to examine this assertion by using sensitivity testing, we would gather a random population of male computational physicists and sort them into age groups (or levels, as age is the explanatory variable). Then we would guess a mean age and choose a number of trials for a Bruceton test series. On either side of the *true mean age*, the numbers of bald and non-bald physicists are distributed evenly. If a computational physicist randomly selected from an age level has hair, the test is a *failure*; if not, the test is a *success*. At the conclusion of the test series, the mean and standard deviation for the normal distribution are estimated (say, in years) by using the methods of Sections 2.4 and 2.5. Of course, there is no stimulus involved in this test, so it does not strictly constitute a sensitivity test, but it is illustrative of our point.

Now consider a similar experiment that involves a more traditional sampling procedure. In this case, we gather a population of "bald", male computational physicists

and decide on a number of sampling trials. We conduct the test series by randomly selecting one physicist at a time and recording his age (in years). After completing the last trial, equations (2.6.3.1) and (2.6.3.2) are used to compute the mean and variance for this distribution. The statistical distribution rendered by this procedure differs from that resulting from the Bruceton procedure. The sampling procedure simply determines the distribution of age among “bald” physicists. These distributions are related, *but they are not identical*. In either case, we can determine the probability that male computational physicists are bald within a certain age group.

By realizing the difference between the data obtained from these two different tests, we can apply the sampling formula for the mean to the sensitivity test series. There are two ways to implement this idea. First, we may perform a direct application of (2.6.3.9) where $\bar{x}$ is the sample mean for the age of the physicists tested. At first glance, this mean seems to have no correlation to hairlessness since it is merely the mean age extracted for the test subjects. One must remember, however, that we are conducting a sensitivity test, so the age levels are designed to reveal the mean based upon the hair/no-hair results. As a result, the value of $\bar{x}$ will have meaning in this context. In the second interpretation of $\bar{x}$, we notice that in analyzing the Bruceton test data, we have estimated values $\hat{\mu}$ for the mean and σ for the standard deviation. We expect that this estimate should, in some sense, coincide with the sample mean. For sensitivity tests, we also formulate our interval estimate based upon binary trials and set n equal one in (2.6.3.9).

2.6.4 Confidence Interval for the Variance

The concept of a confidence interval can also be extended to estimates of the variance. Once again, we imagine repeating an entire series of sensitivity tests over and over again. Remember that the variance is a fixed, but unknown, statistical parameter. Based upon the data, we can estimate both the variance and its confidence interval. As we stated earlier, the endpoints of the interval are random variables since they vary from test series to test series. It should also be stated that, as in the case of the mean, our discussion of the variance confidence interval is based upon sampling formulas for the normal distribution.

Consider the normal distribution $N(\mu,\sigma)$; further suppose that the mean μ is known with a high degree of confidence. The maximum likelihood estimation formula for the variance is

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 , \qquad (2.6.4.1)$$

where X_i is the i^{th} random sample of n drawn from the distribution.[13] Secondly, define the random variable Y such that

$$Y = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 . \qquad (2.6.4.2)$$

Y is distributed as $\chi^2(y;n)$, the chi-square random variable with n degrees of freedom.[12] According to Craig and Hogg, for a $100(1-\alpha)\%$ confidence interval, we seek interval endpoints a and b such that

$$\mathrm{Prob}(a < Y < b) = 1-\alpha , \qquad (2.6.4.3)$$

or by using (2.6.4.2),

$$\text{Prob}\left[a \;<\; \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i-\mu\right)^2 \;<\; b \right] \;=\; 1-\alpha\,. \qquad (2.6.4.4)$$

An algebraic rearrangement of the inequality yields the equivalent probability

$$\text{Prob}\left[\frac{1}{b}\sum_{i=1}^{n}\left(X_i-\mu\right)^2 \;<\; \sigma^2 \;<\; \frac{1}{a}\sum_{i=1}^{n}\left(X_i-\mu\right)^2 \right] \;=\; 1-\alpha\,. \qquad (2.6.4.5)$$

This probability relationship deserves comment. Note that, in general, the interval endpoints are not symmetric, i.e., $a \neq b$. The reason for the difference between a and b is that $\chi^2(y;n)$ is not a symmetric density function; as a result, the choice of these endpoints is *not unique*. As a matter of general practice, a and b are chosen to satisfy the following relationships:[12]

$$\int_0^a \chi^2\left(y;n\right)dy \;=\; \frac{\alpha}{2}; \qquad \int_b^\infty \chi^2\left(y;n\right)dy \;=\; 1-\int_0^b \chi^2\left(y;n\right)dy \;=\; \frac{\alpha}{2}, \qquad (2.6.4.6)$$

where the chi-square density function with n degrees of freedom is defined by[13]

$$\chi^2\left(y;n\right) \;=\; \frac{1}{2^{n/2}\,\Gamma\left(n/2\right)}\,y^{\left(n/2-1\right)}\exp\left(-y/2\right), \quad y \;>\; 0\,. \qquad (2.6.4.7)$$

The presence of the Gamma function in (2.6.4.7) complicates the evaluation of integrals in (2.6.4.6), but with the use of proper numerical algorithms, this process is amenable to digital computation.[15]

The application of interval estimation to the variance presents interesting questions from the standpoint of analyzing sensitivity test data. Similar considerations were posed in the preceding section regarding the estimated mean. Equation (2.6.4.5) really addresses the sample variance attendant to a normally distributed population. Our

sensitivity test results consist solely of go/no-go data, so its connection to classic sampling experiment is less direct. See the preceding section for an example. Nevertheless, (2.6.4.5) and (2.6.4.6) remain valid, but there are two interpretations of the sample variance estimate. In the first interpretation, we directly apply (2.6.4.5); the values of the explanatory variable taken from the test series are used as the X_i while μ is taken equal to $\hat{\mu}$, the estimate of the mean generated by the methods discussed in Sections 2.4 and 2.5. In the second interpretation, we employ mathematical sleight of hand by noting that the sample variance is defined by (2.6.4.1); hence,

$$\sum_{i=1}^{n}\left(X_i-\mu\right)^2 \;=\; nS^2\,. \qquad (2.6.4.8)$$

Garwood's method provides an estimate of σ^2, so we may use this value, denoted $\hat{\sigma}^2$, in lieu of S^2. The confidence interval then is defined by

$$\mathrm{Prob}\left[\frac{n\hat{\sigma}^2}{b} \;<\; \sigma^2 \;<\; \frac{n\hat{\sigma}^2}{a}\right] \;=\; 1-\alpha\,, \qquad (2.6.4.9)$$

with a and b defined as in (2.6.4.6). For sensitivity tests, we apply the analogy for binary trials and set n equal one.

2.7 Measures of Variance

In this section, we address measures of variance associated with our fitting procedure. As it happens, both the normal mean and variance, as statistical estimates, have associated variances Var(μ) and Var(σ), respectively. Also, we can imagine that μ and σ vary jointly across the statistical distribution justifying a calculation of their

covariance. As it happens, these estimates of variance are given as entries in A_{var} the asymptotic variance-covariance matrix, the inverse of the information matrix, A_{info} where

$$A^{\text{var}} = A_{\text{info}}^{-1} = -E\begin{bmatrix} \dfrac{\partial^2 L}{\partial \mu^2} & \dfrac{\partial^2 L}{\partial \mu \partial \sigma} \\ \dfrac{\partial^2 L}{\partial \mu \partial \sigma} & \dfrac{\partial^2 L}{\partial \sigma^2} \end{bmatrix}^{-1}. \tag{2.7.1}$$

A common approach to the determination of the information matrix is to represent each partial derivative as a sum over binary trials.[6] This process requires equations similar to those derived in Section 2.4. It is easy to compute the expected values for these expressions, but we propose an alternative method that uses information already calculated as a part of the solution procedure.

Recall that as a part of our solution algorithm, we defined the Hessian matrix H as

$$H = \begin{bmatrix} \dfrac{\partial^2 L}{\partial \alpha^2} & \dfrac{\partial^2 L}{\partial \alpha \partial \beta} \\ \dfrac{\partial^2 L}{\partial \alpha \partial \beta} & \dfrac{\partial^2 L}{\partial \beta^2} \end{bmatrix}. \tag{2.7.2}$$

The transformation (2.3.2) has the general form of

$$\alpha = \alpha(\mu, \sigma); \quad \beta = \beta(\mu, \sigma). \tag{2.7.3}$$

By using the chain rule for partial differentiation, we can derive the information matrix A_{info} from H. Observe that

$$\frac{\partial}{\partial \mu} = \frac{\partial \alpha}{\partial \mu}\frac{\partial}{\partial \alpha} + \frac{\partial \beta}{\partial \mu}\frac{\partial}{\partial \beta}; \tag{2.7.4}$$

$$\frac{\partial}{\partial \sigma} = \frac{\partial \alpha}{\partial \sigma}\frac{\partial}{\partial \alpha} + \frac{\partial \beta}{\partial \sigma}\frac{\partial}{\partial \beta}. \tag{2.7.5}$$

It is a tedious mathematical exercise, but repeated applications of (2.7.4) and (2.7.5) can be used to show that

$$\frac{\partial^2 L}{\partial \mu^2} = \left[\frac{\partial \alpha}{\partial \mu}\frac{\partial^2 L}{\partial \alpha^2} + \frac{\partial \beta}{\partial \mu}\frac{\partial^2 L}{\partial \alpha \partial \beta}\right]\frac{\partial \alpha}{\partial \mu} + \frac{\partial L}{\partial \alpha}\frac{\partial^2 \alpha}{\partial \mu^2} + \left[\frac{\partial \alpha}{\partial \mu}\frac{\partial^2 L}{\partial \alpha \partial \beta} + \frac{\partial \beta}{\partial \mu}\frac{\partial^2 L}{\partial \beta^2}\right]\frac{\partial \beta}{\partial \mu} + \frac{\partial L}{\partial \beta}\frac{\partial^2 \beta}{\partial \mu^2}. \tag{2.7.6}$$

$$\frac{\partial^2 L}{\partial \mu \partial \sigma} = \left[\frac{\partial \alpha}{\partial \sigma}\frac{\partial^2 L}{\partial \alpha^2} + \frac{\partial \beta}{\partial \sigma}\frac{\partial^2 L}{\partial \alpha \partial \beta}\right]\frac{\partial \alpha}{\partial \mu} + \frac{\partial L}{\partial \alpha}\frac{\partial^2 \alpha}{\partial \mu \partial \sigma} + \left[\frac{\partial \alpha}{\partial \sigma}\frac{\partial^2 L}{\partial \alpha \partial \beta} + \frac{\partial \beta}{\partial \sigma}\frac{\partial^2 L}{\partial \beta^2}\right]\frac{\partial \beta}{\partial \mu} + \frac{\partial L}{\partial \beta}\frac{\partial^2 \beta}{\partial \mu \partial \sigma}. \tag{2.7.7}$$

$$\frac{\partial^2 L}{\partial \sigma^2} = \left[\frac{\partial \alpha}{\partial \sigma}\frac{\partial^2 L}{\partial \alpha^2} + \frac{\partial \beta}{\partial \sigma}\frac{\partial^2 L}{\partial \alpha \partial \beta}\right]\frac{\partial \alpha}{\partial \sigma} + \frac{\partial L}{\partial \alpha}\frac{\partial^2 \alpha}{\partial \sigma^2} + \left[\frac{\partial \alpha}{\partial \sigma}\frac{\partial^2 L}{\partial \alpha \partial \beta} + \frac{\partial \beta}{\partial \sigma}\frac{\partial^2 L}{\partial \beta^2}\right]\frac{\partial \beta}{\partial \sigma} + \frac{\partial L}{\partial \beta}\frac{\partial^2 \beta}{\partial \sigma^2}. \tag{2.7.8}$$

Equations (2.7.6) through (2.7.8) are the elements of A_{info}. We can calculate them easily with the use of the following coordinate transformation derivatives.

$$\frac{\partial \alpha}{\partial \mu} = -\frac{1}{\sigma}; \quad \frac{\partial \alpha}{\partial \sigma} = \frac{\mu}{\sigma^2};$$

$$\frac{\partial \beta}{\partial \mu} = 0; \quad \frac{\partial \beta}{\partial \sigma} = -\frac{1}{\sigma^2};$$

$$\frac{\partial^2 \alpha}{\partial \mu^2} = 0; \quad \frac{\partial^2 \alpha}{\partial \mu \partial \sigma} = \frac{1}{\sigma^2}; \quad \frac{\partial^2 \alpha}{\partial \sigma^2} = -\frac{2\mu}{\sigma^3};$$

$$\frac{\partial^2 \beta}{\partial \mu^2} = 0; \quad \frac{\partial^2 \beta}{\partial \mu \partial \sigma} = 0; \quad \frac{\partial^2 \beta}{\partial \sigma^2} = \frac{2}{\sigma^3}.$$

By inverting A_{info} in accordance with (2.7.1), we obtain an estimate of the asymptotic

variance-covariance matrix; its elements are

$$\mathrm{Var}(\mu) = A^{\mathrm{var}}_{1,1}; \quad \mathrm{Cov}(\mu,\sigma) = A^{\mathrm{var}}_{1,2}; \quad \mathrm{Var}(\sigma^2) = A^{\mathrm{var}}_{2,2}. \tag{2.7.9}$$

3 RESULTS FOR TEST PROBLEMS

In the preceding sections, we have conveyed a part of the probabilistic theory supporting the analysis of sensitivity test data. The mathematics employed is quite complicated and much time is required in order to gain a working level of knowledge in this discipline. Mired in theory, it is easy to lose touch with the practical aspects of the calculations. For this reason, this part of the report presents the setup and results associated with a series of basic test problems. To enhance understanding and permit brevity, we have selected problems from Garwood.[9] These problems are useful in validating the algorithms discussed in this report. Example calculations are provided to illustrate most the analyses described in Section 2. Only the variance confidence interval computation is not shown because of the time required in order to program the Gamma function.

3.1 Example 1 – Antibiotic Efficacy

Our first example addresses a series of serum testing trials associated with an anti-pneumonia drug. Selected doses of the drug are administered to groups of mice. The data is presented as a series of five binomial trials, but the data can also be represented in terms of a sequence of binary trials (each treated mouse constitutes a binary trial). The data is provided in Table 1, and the drug dosage is measured in cubic centimeters.[9] Each individual dosage X_D is administered to 40 mice, and after being infected with pneumonia, the number of expired mice is counted (an expired mouse is denoted as a success in this context). For convenience, the real dosage is mapped onto

Table 1. Data for Example 1 - Testing an anti-pneumonia drug.

Dosage (X_D, mg)	x	Responses out of 40
0.000625	-2	33
0.00125	-1	22
0.0025	0	8
0.005	1	5
0.01	2	2

Figure 1. Probability density (2a) and cumulative distribution (2b) functions for example problem 1. The explanatory variable is serum dosage measured in cubic centimeters (cm^3).

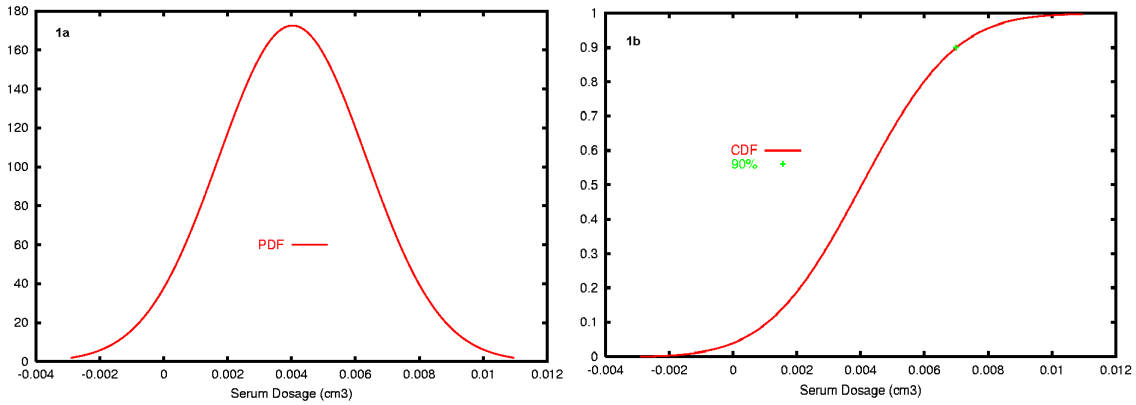


Table 2. Selected non-survival probabilities versus serum dosage for example problem 1.

Dosage (cm^3)	Survival Probability
0.001069	0.1
0.002086	0.2
0.002819	0.3
0.003445	0.4
0.004031	0.5
0.004616	0.6
0.005243	0.7
0.005976	0.8
0.006992	0.9

the interval [-2,2] in accordance with the transformation

$$x = \frac{6}{\max(X_D) - \min(X_D)}\left(X_D - \min(X_D)\right) - 3\,. \qquad (3.1.1)$$

Table 3. 95% confidence intervals for the dosage level versus non-survival probability. Dosage is in cubic centimeters. Dashed endpoint entries "-" indicate the calculation of a negative dosage level. In these cases, the actual endpoint is not defined by the probability model.

Survival Probability	Low Endpoint (cm^3)	High Endpoint (cm^3)
0.1	-	0.0062572
0.2	-	0.0065969
0.3	-	0.0069493
0.4	-	0.0073238
0.5	0.0003296	0.0077326
0.6	0.0007384	0.0081954
0.7	0.0011129	0.0087460
0.8	0.0014653	0.0094490
0.9	0.0018050	0.0105265

This transformation helps control the magnitudes of the partial derivatives needed by the fitting algorithms. It also provides some ease in selecting starting values for μ and σ. As we hinted above, the response is the number of expired mice at the conclusion of the waiting period after the trial. To initiate our estimation procedure, we use the starting values

$$\alpha = 2; \quad \beta = 1, \tag{3.1.2}$$

in equation (2.3.1). The alternation numerical scheme converges quickly yielding the final values of

$$\alpha = 0.554427; \quad \beta = 0.676080. \tag{3.1.3}$$

These values show excellent agreement with Garwood's solution.[9] Moreover, by using the transformations shown in (2.3.2) and (3.1.1), we find the dosage mean and standard deviation in cubic centimeters, i.e.,

$$\mu = 4.03115 \times 10^{-3}; \quad \sigma = 2.31111 \times 10^{-3}. \tag{3.1.4}$$

Table 4. Elements of the asymptotic variance-covariance matrix for example problem 1. Shown are the variances for the normal mean and variance along with the covariance for the normal mean and variance. Units are square cubic-centimeters.

$\mathrm{Var}(\mu)$	$\mathrm{Cov}(\mu,\sigma^2)$	$\mathrm{Var}(\sigma^2)$
6.75×10^{-8}	3.92×10^{-11}	2.18×10^{-13}

Figure (1a) and (1b), respectively, contain plots of the probability density and cumulative distribution functions for this example. The 90% probability is indicated by "+" in Figure (1b). Table 2 contains a list of serum dosage values versus non-survival probabilities based upon our analysis of the experimental data. We have also estimated a 90% confidence interval for the mean as

$$[0.0022971,\ 0.0078325]. \qquad (3.1.5)$$

To provide information on the sensitivity of the dosage estimates, 95% confidence intervals have been calculated for the dosage versus non-survival probability. These estimates are contained in Table 3. The normal distribution is a continuous function defined over the domain $(-\infty, +\infty)$, so in terms of confidence intervals, it is possible to calculate negatively valued dosage endpoints. These endpoints are indicated by dashes in Table 3 since a negative dosage has no physical meaning. From the standpoint of statistics, these interval endpoints are undefined. To obtain the confidence intervals reflected in Tables 2 and 3, we have adapted equations (2.6.2.7), (2.6.2.8) and (2.6.3.9) for binary trials by setting n equal to unity. The resulting intervals seem to concur well with the interval widths predicted by Langlie's Monte Carlo simulations.[16] In addition, the variances of the normal mean and variance are listed in Table 4 along with the covariance of these two parameters.

Table 5. Test series data for example 2. Binomial trials recorded for the number of brine shrimp surviving specified liquid solutions containing arsenic. Arsenic concentrations are in geometrical progression.

Solution	*x*	Responses out of 8
C	-3	8
D	-2	8
E	-1	6
F	0	5
G	1	5
H	2	1
I	3	0

3.2 Example 2 – Arsenic Toxicity

The second test problem examines the susceptibility of brine shrimp to arsenic in liquid solution. As the concentration of arsenic is increased, we expect fewer shrimp to survive. The data for this test series is provided in Table 5.[9] In this case, a response is defined as a shrimp surviving immersion in the solution. Our starting guess for the estimation routine is given by (3.2.1), the same values as used in the preceding example problem. The converged values are

$$\alpha = -0.434164; \quad \beta = 0.712833. \tag{3.2.1}$$

These values constitute an excellent match for the corresponding values in Garwood.[9] The associated normal mean and variance are calculated to be

$$\mu = 0.609069; \quad \sigma = 1.402852. \tag{3.2.2}$$

Since there are no actual values specified for the solution concentrations, these parameter estimates are dimensionless (the same as the dimensions of the explanatory variable *x*). Probability density and cumulative distribution functions are easily calculated for this example problem. These functions are shown in Figures 2a and 2b,

Figure 2. Probability density (2a) and cumulative distribution (2b) functions for example problem 2. The explanatory variable is dimensionless serum dosage.

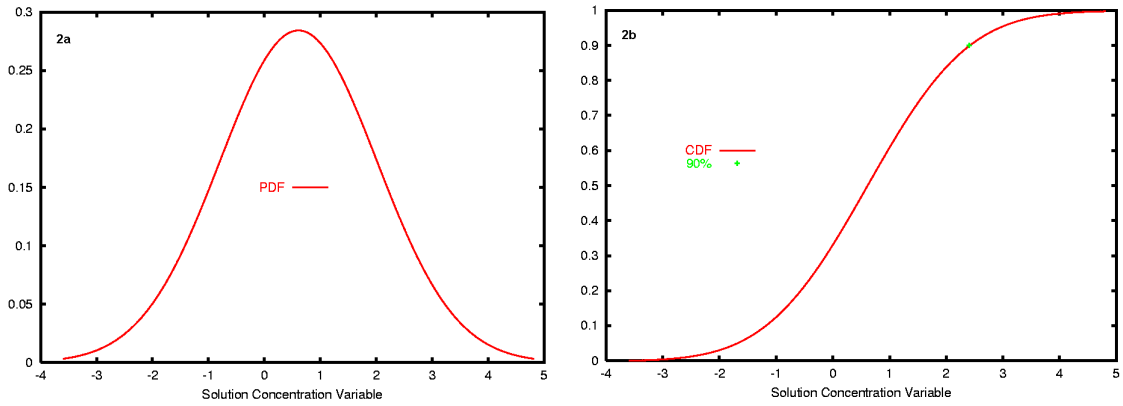


Table 6. Selected survival probabilities versus dimensionless serum dosage for example problem 2.

x	Survival Probability
-1.18876	0.1
-0.57160	0.2
-0.12658	0.3
0.25366	0.4
0.60906	0.5
0.96447	0.6
1.34472	0.7
1.78974	0.8
2.40689	0.9

Table 7. 95% confidence intervals for the arsenic concentration level versus survival probability. Arsenic concentration *x* is dimensionless. Negative values of *x* are allowed.

Survival Probability	Low Endpoint	High Endpoint
0.1	-3.33209	1.96030
0.2	-2.67947	2.16646
0.3	-2.25114	2.38042
0.4	-1.91821	2.60773
0.5	-1.63775	2.85589
0.6	-1.38959	3.13635
0.7	-1.16228	3.46928
0.8	-0.94832	3.89761
0.9	-0.74216	4.55024

Table 8. Elements of the asymptotic variance-covariance matrix for example problem 2, the variances for the normal mean and variance along with the covariance for the normal mean and variance. Variance values are dimensionless.

$\mathrm{Var}(\mu)$	$\mathrm{Cov}(\mu,\sigma^2)$	$\mathrm{Var}(\sigma^2)$
0.10008	0.00967	0.10238

respectively.The point corresponding to a probability of 0.9 is indicated by a "+" sign on Figure 2b. Selected survival probabilities are listed versus the associated value of the explanatory variable *x* in Table 6. Also, the 90% confidence interval for the normal mean is calculated as

$$[-1.69841,\ 2.91655]. \qquad (3.2.3)$$

Since the explanatory variable has no real concentration units, we have allowed negative values for the lower interval endpoint. Table 7 contains the 95% confidence intervals in terms of *x* for a series of probabilities. Finally, Table 8 contains the contents of the information matrix (or asymptotic variance-covariance matrix).

4 SUMMARY

In this report, we have presented a discussion of sensitivity (or Go/No-Go) testing from basic principles. Our prototypical set of sensitivity experiments is based upon the Bruceton test procedure, or “Up and Down Method”, used to estimate the statistical mean associated with a distribution of successful or failed trials. For the “Probit” method, the mean and standard deviation are estimated by fitting data to a normal distribution. The behavior of the normal curve is then used to determine the probability of success associated with an explanatory variable of choice.

The set-up and execution of the Bruceton test procedure have been discussed for an example scenario. We have highlighted the importance of locating the mean and ensuring that extraneous data far removed from the mean is excluded from the analysis. Moreover, issues surrounding the efficiency of the data have been discussed. That is to say, we must obtain a sufficiently large sample in order for our analysis techniques and assumptions to apply.

Algorithms for data analysis have been discussed in detail, particularly the topic on Probit analysis. We have shown the set of equations required to fit go-no go data to the normal curve. This method has proven to be capable and readily addresses non-uniformly separated binomial trials. Although our solution scheme is based upon Garwood’s method, we have presented an alternative to this method that can easily incorporate both binary and binomial trials. The alternate algorithm is stable and remains well-defined in every case. From the standpoint of sampling, we have discussed interval estimation for sensitivity tests. Algorithms have been suggested for the qualified estimation of probability confidence intervals as well as those for the mean

and variance. We have also derived equations that allow the variance and covariance to be estimated for the normal mean and variance produced by our fitting procedure.

Two classic test problems have been chosen from biological science to serve as practical examples. Both problems have been solved by coding developed directly from the algorithms presented in this report. In each case, our results have achieved excellent agreement with archival solutions. We have also produced additional data for these cases through the calculation of confidence intervals. We have also discussed shortfalls that exist within our process for estimating intervals and for estimating elements in the asymptotic variance-covariance matrix.

Sensitivity testing is very important, not only to the Department of Defense, but also to a variety of endeavors in the commercial sector, namely the pharmaceutical industry. The efficacies and effective dosages for medications are largely determined through this type of testing and analysis. In most cases, the statistical distributions produced by the analysis are used to make important management decisions. The investment in the development of a single medication can cost hundreds of millions of dollars. The cost of qualifying explosive systems is also high. For these reasons, it is important that we maintain a good understanding of the probabilistic theory that undergirds sensitivity testing and analysis. System reliability is of paramount importance for protecting both the investment of funding and human life. Failing to accurately estimate the reliability of a modern drug or engineering system can have grave consequences.

REFERENCES

1. Collett, D., Modelling Binary Data, 2nd Ed., *Texts in Statistical Science*, Chapman & Hall/CRC, 2003.

2. Ayres, J.N., Hampton, L.D., Kabik, I. and Solem, A.D., "VARICOMP: A Method for Determining Detonation-Transfer Probabilities", NAVWEPS Report 7411, AD263381, Explosives Research Department, U.S. Naval Ordnance Laboratory, White Oak, MD, July 1961.

3. Crow, E.L., Davis, F.A. and Maxfield, M.W., "Statistics Manual with Examples Taken from Ordnance Development", NAVORD Report 3369, Research Department, U.S. Naval Ordnance Test Station, September 1955.

4. Dixon, W.J. and Mood, A.M., "A method for obtaining and analyzing sensitivity data", *Journal of the American Statistical Association*, Vol. 43, 1948, pp. 109-126.

5. Bliss, C.I., "The calculation of the dosage-mortality curve", *Annals of Applied Biology*, Vol. 22, 1935, pp. 134-167.

6. Golub, A. and Grubbs, F.E., "Analysis of sensitivity experiments when the levels of stimulus cannot be controlled", *American Statistical Association Journal*, Vol. 57, 1956, pp. 257-265.

7. Banerjee, K.S., "On the Efficiency of Sensitivity Experiments Analyzed by the Maximum Likelihood Estimation Procedure Under the Cumulative Normal Response", Technical Report ARBRL-TR-02269, ADA092349, US Army Armament Research and Development Command, Ballistic Research Laboratory, Aberdeen Proving Ground, MD, September 1980.

8. Spahn, P.F. and Montesi, L.J., "Detonation Transfer Reliability: VARIDRIVE and VARICOMP Using Energy Fluence and the Logistic Analysis", IHTR 2292, Indian Head Division, Naval Surface Warfare Center, Indian Head, MD, September 2000.

9. Garwood, F., "The application of maximum likelihood to dosage-mortality curves", *Biometrika*, Vol. 23, 1941, pp. 46-58.

10. Ross, S.M., Introduction to Probability Models, 4th Ed., Academic Press, San Diego, CA, 1989.

11. Cornfield, J. and Mantel, N., "Some new aspects of the application of maximum likelihood to the calculation of the dosage response curve", *American Statistical Association Journal*, Vol. 45, 1950, pp. 181-210.

12. Hogg, R.V. and Craig, A.T., Introduction to Mathematical Statistics, 3rd Ed., Macmillan Publishing, New York, 1970.

13. Larsen, R.J. and Marx, M.L., A Introduction to Mathematical Statistics and its Applications. Prentice-Hall, New Jersey, 1981.

14. Hawking, S., The Universe in a Nutshell. Bantam Books, New York, 2001.

15. Lanczos, C., "A precision approximation of the gamma function", *J. SIAM Numer. Anal.*, Ser. B, Vol. 1, pp. 89-95, 1964.

16. Langlie, H.J., "A Reliability Test Method for One-Shot Items", Report ADP014612, Proceedings of the Eighth Conference on the Design of Experiments in Army Research Development and Testing, 1965.

DISTRIBUTION LIST
AFRL-RW-EG-TR-2009-7005

DEFENSE TECHNICAL INFORMATION CENTER - 1 Electronic Copy (1 File & 1 Format)
ATTN: DTIC-OCA (ACQUISITION)
8725 JOHN J. KINGMAN ROAD, SUITE 0944
FT. BELVOIR VA 22060-6218

AFRL/RWOC-1 (STINFO Office) - 1 Hard (Color) Copy
AFRL/RW CA-N - STINFO Officer Provides Notice of Publication

IIT RESEARCH INSTITUTE/GACIAC - 1 Hard (Color) Copy
10 WEST 35TH STREET
CHICAGO IL 60616-3799