Chapter 1

INTRODUCTION AND LITERATURE REVIEW

§ 1.1 Motivation

The use of statistical methods in the field of engineering is hardly unusual. Most, if not all engineers, are normally required to take one to two semesters of introductory probability and statistical inference. With this statistical foundation, the engineer should be able to recognize basic discrete and continuous distributions (e.g. binomial and normal), apply classical estimation techniques (e.g. sample mean and sample variance), and perform classical statistical hypothesis testing (e.g. Student's t-test and analysis of variance). Unfortunately, while these basic tools are quite useful and widely used, they may not be appropriate for a given set of data. Classical estimation techniques and hypothesis testing require that the data follow a normal distribution and while this is emphasized in introductory statistics, few methods are taught that apply when data deviate from normality. And in the field of engineering, it is customary to find data that does not conform to normality. As an example, there are many possible sources of error with measurements obtained from a power network, mostly due to the violation of some assumptions. Specifically, power engineers assume that the instruments are calibrated, the errors are independent N(0, σ_i^2), and the parameters of the lines are known exactly. In addition, harsh weather and/or electromagnetic interference play a role in affecting the measurements. Obviously, these errors can propagate and affect the statistical estimation.

When dealing with non-normal data, there are a couple of approaches that can be used for estimation and hypothesis testing. One approach is transforming the data to appear normal, so that classical estimation and hypothesis testing are valid. Common transformations include taking a square root or natural logarithm, or using the Box-Cox transformation method (1964). A second approach would be to use nonparametric procedures such as the Wilcoxon Rank Sum test (1945) or the Kruskal-Wallis test (1952). Some nonparametric procedures are based on ranks, thus are not grossly affected by outlying data or skewness in the distribution, as are the classical procedures. A third approach is the use of *robust* procedures, which are much less sensitive to deviations from normality, especially with regards to the tail behavior of the distribution. Common robust estimation techniques include M-estimators based on Huber's function (1964) and Tukey's bisquare function (1974). In the field of electrical engineering, there is a growing interest in both nonparametric procedures and robust procedures for the purpose of estimation and hypothesis testing due to the nature of the data. For example, the use of Tukey's bisquare function for the process of glint noise (Hewer *et al.*, 1987), and the use of Friedman's Two Way Analysis of Ranks for voltage calibration (Ghassemian 1997).

§ 1.2 Statement of the Problem

This dissertation will be comprised of two distinct problems, both with specific applications in power systems engineering. The first is with regard to topology error identification and the second is with regard to voltage calibration. In power systems engineering, topology error identification is a well-documented problem that has been studied over the past twenty-five years. Topology errors occur in power systems and affect power system state estimation. Schweppe and Wildes first derived power system state estimation in 1970. Specifically, they derived a mathematical model of a power network for the power and voltage measurements as a function of the nodal voltages and phase angles. In static state estimation, it is desired to estimate the nodal voltages and phase angles in a given network and this is usually achieved via a weighted least squares algorithm in a nonlinear regression model. However, to properly estimate these nodal voltages and phase angles, it is assumed that the topology of the network is completely known. That is, the status of each line, whether on or off, is known *without error*. This is due to the model matrix being a function of which lines are on and which are off. If a line is on, then it is represented in the nonlinear model; if it is off, it is not represented. If a line is deemed incorrectly as on as opposed to off, or vice versa, a *topology error* occurs which induces an error in the nonlinear function. An error in the latter then translates into bad estimates of these nodal voltages and phase angles, if the algorithm converges. Since this is a nonlinear regression model, the estimates must be solved iteratively, and topology errors may cause the estimator to diverge. Thus, a robust method of detecting topology errors before any state estimation is desired.

The Friedman Two Way Analysis of Variance by Ranks test (Friedman, 1937) is a nonparametric test advocated for voltage calibration, due to the noisy nature of the data. In this scenario, each meter (assume there are *t* meters) at a bus records a voltage measurement, and this is done periodically over a small interval of time (assume *b* periods of time). This is a randomized complete block setting where the meters are the treatments and the time periods are the blocks. With the voltage measurements as responses, the Friedman test is used to test the null hypothesis of no treatment effects, where failing to reject the null hypothesis is the desired conclusion. If the null is rejected, then post-hoc tests are done to see which of the meters do statistically agree. However, Ghassemian (1997) noted that in the presence of bad data, the Friedman test was only slightly more robust than the classical F-test. He saw that with a few bad data, the test became unreliable. This prompts the question of exactly how robust or *resistant* the Friedman test is in the presence of outlying data.

§ 1.3 Review of Topology Error Identification

As mentioned previously, Schweppe and Wildes derived power system state estimation in 1970. Commonly, weighted least squares estimation is employed to estimate the state, and post-analysis tests on the normalized residuals are used to detect and eliminate gross errors in the data. Assuming that the topology of the network is correct, any high residuals are due to measurement errors. If the topology of the network is not correct, then multiple conforming bad data will emerge and cause inaccurate state estimates, if the weighted least squares algorithm converges. By 'conforming', we mean that the measurements on the line where the topology error has occurred, as well as on adjacent busses and lines, have 'shifted' in the design space. Robust estimators based on the ℓ_1-norm, as shown by Clements *et al*. (1991) and Çelik and Abur (1991), the Schweppe type Huber GM estimator as shown by Mili *et al.*(1995), and on the least median of squares, as shown by Mili *et al.* (1991), have improved state estimation in the presence of multiple bad data. However, robust estimators such as the ℓ_1-norm will break down in the presence of leverage points, and estimators such as the LMS can be extremely computationally intensive. Regardless, these estimators do not help with inaccurate estimates of the state due to topological errors that cause clusters of bad data in an area of the network, which is explained in more detail in Chapter 2. The topic of topology error identification has been studied intensely over the past decade. Such methods of detection include rule-based procedures (Singh and Oesch, 1994), statistical procedures (Abur *et al.*, 1995), and equality-constraint procedures (Clements and Costa, 1998). The rule-based procedures assess the status of a line or set of lines via associated circuit breaker positions, measurement conformity based on Kirchhoff's Law, and past data. From electrical engineering, Kirchhoff's Law states that the algebraic sum of the total power in the network is zero. That is, the power injected into the network equals the power consumed. Via rule-based procedures, a line's status is determined by a point system based on these rules, and while this scheme is simple, its performance has yet to be studied in earnest detail. Abur *et al.* proposed implementing a weighted least absolute value estimator to reject bad measurements and to detect topology errors. The procedure is two step in nature; the first step detects high normalized residuals and selects a set of suspect faulty lines associated with these high residuals. The second step extends the original state vector and additionally estimates a 'branch flow error' for each suspect branch. The line with highest estimated branch flow error is flagged as a topology error. Unfortunately, what constitutes a 'high' branch error is not addressed, as well as the fact that this approach works for just one type of topology error. Clements and Costa recently introduced a method of topology error identification based on the use of Lagrangian multipliers that are affiliated with every line. A weighted least squares algorithm is employed subject to *operational constraints* and *structural* *constraints*. The Lagrange multipliers are then normalized and tested against a critical threshold with large values indicating a topology error. However, for this approach the state vector not only includes the regular state variables, but the Lagrangian multipliers as well, thereby almost doubling the number of unknowns.

§ 1.4 Review of Test Resistance

Loosely stated, a statistical test that is impervious to bad data is said to be *resistant*. More specifically, the resistance of a statistical test is the number of contaminated observations necessary to switch an acceptance of the null hypothesis to a rejection, or vice versa, divided by the total sample size. Thus, a test's resistance is akin to an estimator's breakdown point (Donoho and Huber, 1983). Ylvisaker (1977) first formally defined a test's resistance in the univariate setting by assuming an addition contamination model. That is, one adds contaminated observations to an already existing sample until finally changing the verdict. However, this may not be as realistic a scenario as a replacement contamination model, in which the data are altered in an already existing sample of size N. Replacement contamination is also commonly used in many other definitions of robustness. For example, Donoho and Huber (1983) used this model for their definition of the finite sample breakdown point. For other specific examples, see Rousseeuw and Leroy (1987) and Simpson (1989). It is under this model that Coakley and Hettmansperger (1992) defined the maximum resistance to acceptance (*MRA*) and the maximum resistance to rejection (*MRR*), as well as the expected resistance to acceptance (*ERA*) and expected resistance to rejection (*ERR*). The maximum resistance is a measure of how robust a test is when the data are in the least favorable position with regards to a desired conclusion, while the expected resistance is a measure of how robust a test is on the average with regards to a desired conclusion. Coakley and Hettmansperger (1994) derived the maximum resistances of the one sample t-test and sign test, as well as the two sample pooled *t*-test and Mood's test. They also derived the expected resistance to rejection of the sign test and simulated the expected resistance to rejection of six two-sample tests. Under an addition contamination model, Zayed and Quade (1997) compared the resistances of five rank correlations, as well as derived an asymptotic distribution of the resistance to rejection of rank tests (1997). However, the maximum and expected resistances in the two-way layout have yet to be examined.

§ 1.5 Goals

The goals of this dissertation are twofold. First, a new robust procedure for identifying topology errors in a network will be derived and tested. The procedure should be able to discriminate between topology errors and gross measurement errors and be able to handle both simultaneously. Second, the resistances of the Friedman Two Way Analysis of Variance by Ranks test will be derived and simulated. With the Friedman test as a base, another test for the two-way layout will be studied. The maximum and expected resistances will be derived and simulated and comparisons will be made.

This dissertation is divided into these two separate topics with Chapters 2 and 3 addressing topology error identification and Chapters 4-8 addressing statistical resistance. Chapter 2 begins by describing and illustrating a basic power network. This is followed by the concept of state estimation and the problems that occur with state estimation in the presence of topology errors. Topology error identification is then reviewed by examining in some detail three proposed types of identification schemes. In Chapter 3 we develop a new procedure for topology error identification by deriving a statistical model of the real and reactive power flows and injections and by using Kirchhoff's Current Law as a foundation for the network modeling. The new method uses the Huber estimator and IRLS algorithm to estimate the correct topology of the network and simulation results on the IEEE-118 bus network show that the methodology was very powerful at detecting both topology errors as well as gross measurement errors. The dissertation switches topics in Chapter 4 by introducing and reviewing the topic of statistical test resistance by highlighting some definitions and examples by Ylvisaker and by Coakley and Hettmansperger. We formulate in Chapter 5 the maximum resistance to rejection and acceptance of the Friedman test statistic. The exact maximum resistance to rejection is derived for the case where the number of blocks is equal to the number of treatments, and is accurately approximated for the case where the blocks are a positive integer multiple of the number of treatments. The maximum resistance to acceptance is approximated for general case when the number of blocks is an integer multiple of the number of treatments, but is shown that as the number of blocks tends to infinity, the *MRA* tends to ¼. Chapter 6 shows the simulated expected resistance to rejection and to acceptance of the Friedman test statistic and compares them directly to the maximum resistances derived in Chapter 5. In Chapter 7 we explore the Brown-Mood test statistic and derive a lower bound of the maximum resistance to rejection as well as derive the maximum resistance to acceptance. We compare these measures to the derived Friedman maximum resistances from Chapter 5 and show that the Brown-Mood test statistic has a higher maximum resistance to rejection and to acceptance. Chapter 8 displays the simulated expected resistance to rejection and expected resistance to acceptance of the Brown-Mood test and compares these to the Brown-Mood maximum resistances derived in Chapter 7 and to the expected resistances of the Friedman test from Chapter 6. We show that Brown-

Mood test is only slightly more robust on average than the Friedman with regards to rejection, but that the Friedman test is much more robust on average with regards to acceptance. Finally, Chapter 9 finishes with conclusions and future research topics for both of the problems presented in this dissertation.