Order Number 9424199

Signal detection using categorical temporal data

Cannon, Ann C. Russey, Ph.D.

Iowa State University, 1994

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

Signal detection using categorical temporal data

by

Ann C. Russey Cannon

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Statistics
Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1994

TABLE OF CONTENTS

LIST OF FIGURES

1. GENERAL INTRODUCTION

1.1 Motivating Experiment

Over the years, several very effective drugs have been found to have serious side effects, often discovered long after the primary cure has been effected. Some of these side effects might have been detected by behavioral differences in the patient although the side effect is most likely physiological. Researchers have designed experiments using rats as an animal model to try to predict the possibility of a side effect measurable by a behavior change for drugs developed for humans. This dissertation discusses the analysis for data produced from an experiment designed to detect such behavioral differences.

The specific design of the experiment is as follows (e.g. Kernan, Mullenix, Kent, Hopper and Cressie, 1988): pairs of rats were chosen, one in each pair injected with a saline solution, the other with the experimental drug. At some later point in time these pairs were observed as they explored a novel environment and their act or position at the beginning of each of the 900 consecutive seconds was recorded. In earlier experiments, the experimenters used 900 still-frame photographs and visual discrimination to code the photographs; more recently, experimenters have used television cameras and a computer pattern recognition system to do the coding. Observed behavioral differences between the rats provide evidence of physiological change.

The data from this experiment were coded as follows. The behavior of the rat at each of the 900 time points was categorized into one of five body positions: standing, sitting, rearing, walking, and lying down; and one of the eight modifiers: blank (no recognized activity), groom, head turn, turn, look, smell, sniff, and washing face. The combined set of 5 body positions and 8 modifiers are defined to be the set of 13 regular acts. Norton, Mullenix and Culver (1976) reclassified each of the regular acts into one of three classes labeled "grooming," "exploratory," and "attention." Since each of the 900 seconds consists of two regular acts, both a body position and a modifier, Kernan and Meeker (1992) mapped the combination of body position and modifier at each second into one of the six combinations of two classes, e.g. "attention-attention," "attention-explore," "groom-attention." It should be noted that these new combinations did not reflect which classification each specific regular act had separately, merely the two classes that the two regular acts together represented. These six combinations are labeled the combined acts. Kernan and Meeker (1992) then used the information from both the regular acts (of which there are two for each time period) and the combined acts (of which there is one for each time period) in their analysis of the data.

The experiment described above leads to a raw data set with observations on $2n$ rats, where n denotes the number of pairs observed. Each observation consists of the following information: a pair number, a treatment/control indicator variable, and, for each of the 900 discrete time periods, the values of the two categorical response variables that represent the body position and modifier, respectively, of the act that the rat performed. From these data the goal is to detect systematic behavioral differences that exist between the treated and the control rats. At this stage the

amount of difference is not of primary concern, merely whether a difference, not due to usual rat-to-rat variability, exists or not. If a systematic difference is found, the researchers would use this as evidence that some physiological change may have taken place and more testing would be done to find the particular physiological difference in question. At the outset of these experiments, the researchers often do not have an idea about what type of systematic behavioral difference might exist, so they look for any detectable change.

Several ways of quantifying data from this type of experiment have evolved. Originally, researchers simply counted the number of times that each act was initiated (Kernan et al., 1988). Psychiatrists, however, found that the total time a rat performed an act, the average duration of the acts, and the time lapsing between initiations of an act were also of interest (Pohl, 1976; Baumeister, 1978). Wender (1971) also showed that some types of behaviors, such as hyperactivity in children, will not be discovered using only the number of initiations, the total time, the average duration and the time lapsing between initiations. Hyperactivity has been traditionally seen as a state of increased motor activity, but this sense of increase comes from the behavior occurring in a different pattern than normal rather than a true increase in activity level. An unexpected behavior pattern may lead to the appearance of more activity while the average amount of time spent doing each act is really the same as in a normal child (Pontius, 1973). In order to detect a difference in the pattern of the behavior from the expected pattern of behavior, Kernan et al. (1988) looked more into the structure of the observations, using a temporal analogue of the spatial statistic, the K function, to define what they called the time distribution of the acts and the time sequence of pairs of acts.

Kernan, Mullenix and Hopper (1989) went one step further by looking at many test statistics aimed at describing differences between the groups for various aspects of the data, hoping that by looking at many statistics, any possible behavioral difference might be detected. The researchers then based their conclusions on which statistics were found to be significant (Kernan, Mullenix and Hopper, 1989). The Kernan, Mullenix and Hopper (1989) method evolved into a new overall test statistic, consisting of the ratio of the number of individual test statistics that were found to be significant to the number of tests done (Kernan and Meeker, 1992). By using a jackknife-type simulation method with control-control data, Kernan and Meeker (1992) computed the null distribution for this overall statistic for experiments consisting of 17 and 20 pairs of rats.

1.2 Most Recent Data Analysis

Kernan and Meeker (1992) started by computing four types of individual statistics for both the set of combined acts as well as the set of regular acts. These four statistics are:

1. *Number of Initiations.* An initiation of an act was defined to be the first frame of the one or more consecutive frames where the animal performed the specific act.

2. *Total Time.* The total number of frames where the act occurred.

3. *Time Distribution.* A statistic describing the pattern of where an act occurs over time.

4. *Time Sequence of Pairs of Acts.* A statistic describing the pattern of how a pair of acts occurs over time.

These last two types of statistics are based on estimating the temporal analogue to the spatial statistic, the K function. Kernan and Meeker (1992) used an estimator of the K function derived in Kernan et al. (1988) to define the time distribution and an estimator of the cross K funtion to define the time sequence of pairs of acts for the two groups of rats.

1.2.1 K functions

Both the number of initiations of an act and the total time spent performing an act are straightforward to calculate. The time distribution, however, is complicated enough that we explain its background, meaning and use. The K function, sometimes called the reduced second moment measure, gives a quantification of the spatial dependence between different regions of a stationary point process. K functions have generally been used on regions that are of two or higher dimensions. They can, however, be quite useful in one dimension. With a minor modification in definition, the K function can also be used in the (unidirectional) time dimension to characterize temporal dependence in a temporal point process.

The spatial K function, with intensity λ, is defined theoretically in d-dimensional space ($d \geq 1$) as:

$$K(h) \equiv \lambda^{-1} E(\text{number of extra events within distance } h \text{ of a randomly chosen event})$$

(e.g., Diggle, 1983, p.47). The 1-dimensional, unidirectional, temporal K function is

defined as:

$$K(t) \equiv \lambda^{-1} E(\text{number of extra events occurring within time } t \text{ } \textit{after} \text{ an arbitrary event.})$$

These functions are used to identify a structure in the pattern of initiations of an act over time. It is possible to find the same number of initiations and the same amount of time spent performing a specific act for treated animals as for control animals, but find that the pattern of the initiations of this act to be different. For example, one type of animal may have seemingly random initiations of an act where the other might display a nonrandom pattern. One use of K functions is in the calculation of the time distribution of an act. This statistic is used to find any time dependence or differences in time dependence that might exist for the initiations of an act.

The K-function estimator used by Kernan and Meeker (1992) in the time distribution was derived in Kernan et al. (1988) and is written as

$$K(t) = \frac{T}{(n^2)} \left[\sum \sum_{i \neq j} W_{ij}^{-1} I_t(U_{ij}) \right]$$

In this equation n denotes the number of initiations of the act, and T denotes the total time over which the analysis is performed. This time, T, is, however, not the same as the total time that the rats were observed. The K function analysis is performed only on the initiations of the act. The duration of the time that the rats performed the act continuously is not of interest here. Thus we reduce the original data vector somewhat by eliminating the continuations of the act of interest after each of its initiations. T represents the length of this reduced data vector. W_{ij} is an edge correction term which accounts for the fact that the observational period is

not infinite. If the act occurs near the beginning or the end of the observational time period not all ranges of time are available for consideration. The edge correction, W_{ij}, is assessed by Kernan et al. (1988) as follows: First, assume that the ith occurrence of act α happened before the jth. Next, let

$$d_{ij} = |t_i - tj| = |t_j - t_i| = d_{ji}$$

Define $W_{ij} = 1$ unless $t_i - d_{ij} < 0$. In this case define $W_{ij} = .5$. Similarly define $W_{ji} = 1$ unless $t_j + d_{ij} > T$ in which case $W_{ji} = .5$ (Kernan et al., 1988). $I_t(U_{ij}^{\alpha})$ is 1 (or 0) according to whether the pair (i, j) of initiations of the act occurred (did not occur) within a time separation t. Mullenix et al. (1989) evaluated this function at eight time points (2, 5, 10, 20, 30, 45, 100, 200) for each act, with the resulting vector being referred to as the time distribution of the act.

The cross K-function is used by Kernan et al. (1988) to calculate the time sequence of a pair of acts. For this purpose call the acts in question δ and γ. The function is

$$K_{\delta\gamma}(t) = (n_\delta n_\gamma)^{-1} T_{\delta\gamma} \left[\sum_{i=1}^{n_\delta} \sum_{j=1}^{n_\gamma} W_{ij}^{-1} I_t(U_{ij}) \right]$$

The terms of this function have similar definitions to those in the time distribution function. Here $T_{\delta\gamma}$ is corrected for the continuations of both acts δ and γ. For the cross K-function, $I_t(U_{ij}) = 1$ if the ith occurrence of act δ happened within time t before the jth occurrence of act γ. Otherwise $I_t(U_{ij}) = 0$. This definition of $I_t(U_{ij})$ is purposefully asymmetrical, in order to allow a causal relationship to exist (Kernan and Meeker, 1992).

Kernan and Meeker (1992) also placed a restriction on the calculation of the K functions in general in order to limit consideration to those statistics for which there

was sufficient information. Their rule for the calculation of the K functions was to use only those acts that had an average number of initiations per animal of ten or greater for both the control group and the exposed group separately. They believed that acts with less initiations would not bring enough accurate information into the function.

1.2.2 Kernan and Meeker (1992) single test statistic

First each of the four statistics defined at the beginning of Section 1.2 had to be computed for each of the regular and combined acts (or pairs of acts) defined in Section 1.1. Because of the restriction on the use of K functions described above, the body position "lying down" was never used in the computation of the statistics "time distribution" or "time sequence". Therefore there were 13 "total time" statistics, 13 "number of initiations" statistics, a maximum of 12 "time distribution" statistics and a maximum of 68 "time sequence" statistics computed using the 13 regular acts (5 body positions and 8 modifiers) for a maximum of 106 statistics. There were also 6 "total time" statistics, 6 "number of initations" statistics, a maximum of 6 "time distribution" statistics, and a maximum of 30 "time sequence" statistics using the 6 combined acts for a maximum of 48 statistics. The resulting maximum number of statistics for a data set is 154. Some of these statistics may not be calculated for every experiment of this type because of the restrictions mentioned above, but a large majority will be used. At this point, then, there were approximately 154 statistics to use in answering the question "Is there a difference between the two groups?" Kernan and Meeker (1992) suggested two ways of combining the statistics to help answer the question. First they created a vector $\mathbf{X}$, of length at most 154, containing

the statistics that can be used to test for differences between the two groups. Then $\mathbf{X}$ was mapped into a vector $\mathbf{Y}$ of 0's and 1's where a 1 signified that the corresponding element of $\mathbf{X}$ showed a statistically significant difference.

At this point the analysis took into consideration the fact that each observation of the animals was used twice, once in the computation of the statistics for the regular acts and again for the combined acts. Because of this overlap, Kernan and Meeker (1992) divided their vector $\mathbf{Y}$ into two parts, $\mathbf{Y}_1$ based on the regular acts and $\mathbf{Y}_2$ based on the combined acts. They designated S_1 and S_2 as the sums and T_1 and T_2 as the lengths of these two vectors respectively. In their example $T_1 \leq 106$ and $T_2 \leq 48$. The two statistics that Kernan and Meeker (1992) suggested are

$$RTOT = \frac{S_1 + S_2}{T_1 + T_2}$$

and

$$RSQR = \left(\left(\frac{S_1}{T_1} \right)^2 + \left(\frac{S_2}{T_2} \right)^2 \right)^{\frac{1}{2}}.$$

Because of the structure of the actual data analysis, T_1 will always be larger than T_2 which results in $RTOT$ more heavily weighting the information from the regular acts. $RSQR$ weighs each of the subvectors more evenly, possibly a desirable condition since each subvector, in some sense, summarizes the data set.

Kernan and Meeker (1992), using previous control-control and control-treatment data and jackknife techniques, explored the distributional characteristics of both statistics. They found that there was an extremely high degree of correlation between $RSQR$ and $RTOT$, indicating that either could be used. By using Monte Carlo calculations on a data set from a control-treatment experiment where a systematic behavior difference was known to exist, they showed that these statistics can

be useful in determining whether a systematic behavioral difference exists or not. The simulation results presented in Kernan and Meeker (1992), however, only give procedures for experiments with 17 or 20 pairs of animals. They suggest that one might be able to interpolate, using their results, for 18 or 19 pairs, under a similar experimental design, but their work does not directly apply to other numbers of pairs. For experiments of other sizes more simulations would have to be run.

1.3 Dissertation Organization

This dissertation consists of two papers combined into one document. The first paper takes a closer look at the K-function estimator. The K function is typically used in a space which has two or more dimensions. Kernan et. al. (1988) took the existing estimator of the K function and brought it down to the unidirectional time dimension. Since their work, a better estimator has been developed for two and higher dimensions (Stein, 1993). The first paper takes this new estimator for the K function, shows that it is appropriate for the time dimension, and applies it to the problem described above. Specifically, two test statistics based on the estimate of the K function are developed to test for differences between control and treated rats. This paper deals only with the K-function estimator for use in the time distribution of acts. Similar results should hold for the time sequence of pairs of acts using the cross K-function, but proof of this result is left to future work. This paper was written in collaboration with co-major professor Noel Cressie.

The second paper uses the new estimator for the temporal K function along with the statistics total time, and number of initiations to create two new overall methods for analyzing the data from this type of experiment. Kernan and Meeker (1992) in

essence counted the number of significant tests among a large group of tests. The first new statistic is a sum of the squared individual test statistics. This statistic is developed with the idea that it will take into consideration the magnitude of the significance of the individual tests as well as the fact that a test is significant. The second new statistic uses more information yet, namely the covariance matrix of the individual test statistics, to create a Wald-like statistic. It was thought that this second new statistic would be superior to both the old statistic and the first new statistic because of the use of the covariance matrix. This paper was in collaboration with co-major professor William Meeker.

A general conclusion chapter follows the two papers and all references cited in the general introduction and conclusion are listed following the general conclusion.

2. TEMPORAL ANALOGUES TO SPATIAL K FUNCTIONS

A paper submitted to the *Biometrical Journal*

Ann Cannon and Noel Cressie[1]

2.1 Abstract

In this article, the spatial statistic known as the K function is adapted for temporal processes and patterns. The (optimal) K-function estimator is used in a testing procedure to determine whether behavior patterns of exposed rats versus control rats are different. Specifically, the temporal analogue to the K function is given and an approximately optimal estimator is developed. Next, a testing procedure, to determine whether a group of point patterns is generated from complete temporal randomness, is given. Finally, a testing procedure, to compare pairwise two groups of point patterns, is given. The testing procedures are illustrated with rat-behavior data from both a control-control experiment as well as an exposed-control experiment, where in the latter case a difference in behavior is known to exist.

[1]Principal Author is Ann Cannon. Noel Cressie provided guidance throughout the project

2.2 Introduction

The K function, sometimes called the reduced second moment measure, gives a quantification of the spatial dependence between different regions of a stationary point process. K functions have generally been used on regions that are of two or higher dimensions. They can, however, be quite useful in one dimension. With a minor modification in definition, the K function can also be used in the (unidirectional) time dimension to characterize temporal dependence in a temporal point process.

The spatial K function, with intensity λ, is defined theoretically in d-dimensional space ($d \geq 1$) as:

$$K(h) \equiv \lambda^{-1} E(\text{number of extra events within distance } h \text{ of a randomly chosen event})$$

(e.g., Diggle, 1983, p.47). The 1-dimensional, unidirectional, temporal K function is defined as:

$$K(t) \equiv \lambda^{-1} E(\text{number of extra events occurring within time } t \textit{ after} \text{ an arbitrary event.}) \tag{2.1}$$

In this article we give both an estimator of the K function suitable for the temporal case, as well as a method for analyzing one or two groups, each consisting of several estimates. We begin (in Section 2.3) by finding the temporal analogue to an initial, well known, d-dimensional K function estimator due to Ohser and Stoyan (1981). We then show how to improve on the initial estimator in the temporal case and give the (approximately) optimal estimator. In Section 2.4 we give a method for comparing the K-function estimates from one sample of data to the null hypothesis of complete temporal randomness. We illustrate this method with rat behavior data

that is expected to follow this null hypothesis. Section 2.5 gives a method for analyzing the difference between two samples of estimates that are paired. In this case we do not necessarily assume complete temporal randomness for either of the two samples. Again we illustrate the use of this method with rat behavior data; the two samples consist of a control group and an exposed group. Section 2.6 contains some final discussion.

2.3 Properties of the Temporal K Function

To find appropriate estimators for the K function, we first obtain several general distributional results. Where a specific distribution is needed, we use the case of complete temporal randomness (CTR) as the basis for our results. The CTR model was chosen because it is the natural null hypothesis from which one determines departures of either a clustered or regular nature. In Sections 2.4 and 2.5, CTR arises naturally in the application presented, where it indicates a lack of pattern in the initiation of various behaviors.

The exact K function for a stationary temporal point process can be calculated as follows. First, let $N(dv)$ denote the number of events occurring in the time span dv, located at v; $v \geq 0$, and consider the following probability:

$$\Pr(\text{there is an event in the time span } du, \text{ located at } u \mid \text{an event at } 0).$$

This probability can be rewritten as:

$$\lim_{d0 \to 0} \frac{\Pr(N(du) > 0, N(d0) > 0)}{\Pr(N(d0) > 0)}. \tag{2.2}$$

A CTR point process is a Poisson process with parameter λ. That is, the number of events occurring in disjoint regions are independent and follow a Poisson distri-

bution. As a consequence, if $N(dv)$ denotes the number of events occurring in the infinitesimal time span dv located at v, then $\Pr(N(dv) > 0) = \lambda dv + o(dv)$ and $\Pr(N(du) > 0, N(dv) > 0) = (\lambda du + o(du))(\lambda dv + o(dv))$. Therefore, under CTR (2.2) becomes: $\lambda\, du$. The K function in (2.1) is defined to be a constant times the expected value of the number of events happening within time t after an event. This expectation can be rewritten as

$$\int_0^t \Pr(\text{there is an event in the time span } du, \text{ located at } u \mid \text{an event at } 0) du.$$

Combining equation (2.1) and equation (2.2) for a Poisson process,

$$K(t) = \lambda^{-1} \int_0^t \lambda\, du = t; \qquad \text{for } t > 0. \tag{2.3}$$

To make inferences about a process, we can use data in the form of an observed point pattern to estimate a K function. Several estimators for the spatial K function are available; some have analogues in time. Time analogues of K function estimators are all based on counting the number of pairs of events located in a time interval less than or equal to t and then averaging this count over the number of points in the observation region. The presence of boundaries (or edges) that limit the observation region means that not all ranges of time are available for consideration. The main difference among estimators is the way that they correct for the edges of the region. One *two-dimensional* estimator due to Ripley (1976) uses a weight equal to the conditional probability that an event is observed given that it is a distance h away from the current reference event. Unfortunately there is no natural analogue to this in the temporal case so we look elsewhere for temporal K-function estimators.

2.3.1 An initial temporal K function estimator

Consider now the two-dimensional estimator suggested by Ohser and Stoyan (1981). Their edge correction is achieved using a weight that is based on the conditional probability of observing a pair of events $(\mathbf{s}_i, \mathbf{s}_j)$ given that they are a distance $|\mathbf{s}_i - \mathbf{s}_j|$ apart. In the temporal case the analogue to the Ohser and Stoyan's edge correction is the conditional probability of observing a pair of events (t_i, t_j), where $t_j > t_i$, given that they are a time lag of $t_j - t_i$ apart. This probability is

$$\frac{T - (t_j - t_i)}{T}.$$

Thus, the temporal analogue to Ohser and Stoyan's two dimensional estimator of the K function is:

$$\hat{K}(t) = \frac{1}{\lambda^2} \sum_{t_j > t_i} \frac{I((t_j - t_i) < t) I(t_j > t_i)}{T - (t_j - t_i)}, \tag{2.4}$$

where the summation is over $\{(t_i, t_j) : t_j > t_i \text{ and } t_i, t_j \epsilon [0, T]\}$.

To calculate the variance of this K-function estimator, we rewrite (2.4) as

$$\hat{K}(t) = \frac{1}{2} \left(\frac{1}{\lambda^2} \right) \sum_{t_i \neq t_j} \phi_t(t_i, t_j). \tag{2.5}$$

where

$$\phi_t(t_i, t_j) = \frac{I(|t_i - t_j| < t)}{T - |t_i - t_j|}. \tag{2.6}$$

These two expressions for the K function are equivalent, the difference being that the summand $\phi_t(\cdot, \cdot)$ in (2.5) is symmetric (i.e. $\phi_t(t_i, t_j) = \phi_t(t_j, t_i)$). The multiplication by $1/2$ corrects for counting each pair of events twice.

To compute the expectation (under stationarity) and variance under the more

specific case of CTR of $\hat{K}(t)$, let

$$B = \sum_{t_i \neq t_j} \phi_t(t_i, t_j).$$

Now, under stationarity,

$$\mathrm{E}(B) = \int_{A^2} \phi_t(t_i, t_j) d\alpha_2(t_i, t_j), \tag{2.7}$$

where α_2 is the second factorial moment measure (e.g., Daley and Vere-Jones, 1972, sec 3.1) for the process, and $A = [0, T]$ denotes the observation region. Similarly,

$$\begin{aligned} \mathrm{E}(B^2) &= \int_{A^4} \phi_t(t_i, t_j)\phi_t(t_k, t_l) d\alpha_4(t_i, t_j, t_k, t_l) \\ &\quad + 4 \int_{A^3} \phi_t(t_i, t_j)\phi_t(t_i, t_k) d\alpha_3(t_i, t_j, t_k) \\ &\quad + 2 \int_{A^2} [\phi_t(t_i, t_j)]^2 d\alpha_2(t_i, t_j) \end{aligned} \tag{2.8}$$

where α_3 and α_4 are the third and fourth factorial moment measures, respectively. Under CTR, $\alpha_r = \lambda^2 \nu_r$, where λ is the rate of the Poisson process and ν_r is the Lebesgue measure in R^r, which leads to

$$\mathrm{var}(B) = \mathrm{E}(B^2) - [\mathrm{E}(B)]^2 = 4\lambda^3 S_1 + 2\lambda^2 S_2 \tag{2.9}$$

where

$$\begin{aligned} S_1 &\equiv \int_{A^3} \phi_t(t_i, t_j)\phi_t(t_i, t_k) dt_i dt_j dt_k \qquad (2.10) \\ &= \int_A \left[\int_A \phi_t(t_i, t_j) dt_j \right]^2 dt_i \\ S_2 &\equiv \int_{A^2} (\phi_t(t_i, t_j))^2 dt_i dt_j. \qquad (2.11) \end{aligned}$$

Because $\hat{K}(t) = \frac{1}{2\lambda^2} B$, we have that $\mathrm{var}(\hat{K}(t)) = \frac{1}{4\lambda^4} \mathrm{var}(B)$.

Details of the derivations of the following results can be found in Appendix A. For $\phi_t(\cdot,\cdot)$ given by equation (2.6), $\hat{K}(t)$ is unbiased for the K function and

$$\text{var}(\hat{K}(t)) = \frac{4(T-t)}{\lambda}\log\left(\frac{T-t}{T}\right) + \frac{4t}{\lambda} + \frac{1}{\lambda^2}\log\left(\frac{T}{T-t}\right). \qquad (2.12)$$

Figure 2.1 shows a plot of this variance as a function of t. The plot shows the variance increasing with t, which is true in general because

$$\frac{\partial}{\partial t}\text{var}(\hat{K}(t)) = \frac{4}{\lambda}\log\left(\frac{T}{T-t}\right) + \frac{1}{\lambda^2(T-t)} \geq 0, \qquad \text{for } t < T.$$

To obtain an estimator that can be used with equal confidence across all values of t under consideration, we shall attempt to find a function g such that the variance of $g(\hat{K}(t))$ is approximately constant across time.

It was shown earlier that $\text{E}(\hat{K}(t)) = t$ and the variance was given by eq. (2.12). Let c be a constant and let V_λ represent the variance function (2.12). An application of the δ-method (e.g., Bishop, Feinberg, and Holland, 1975, p. 491) to $\text{var}(g(\hat{K}(t)))$ yields the following condition:

$$\left[\frac{\partial}{\partial y}g(y)\Big|_{y=(\text{E}(\hat{K}(t)))}\right]^2 V_\lambda(t) = c.$$

To find g solve

$$g(y) = \int_0^y \frac{c}{(V_\lambda(t))^{1/2}}dt.$$

A Taylor series expansion can be used to approximate $V_\lambda(t)$. The first two terms of this approximation are:

$$V_\lambda(t) \approx \frac{1}{\lambda^2 T}t + \left(\frac{1+4\lambda T}{\lambda^2 T^2}\right)t^2.$$

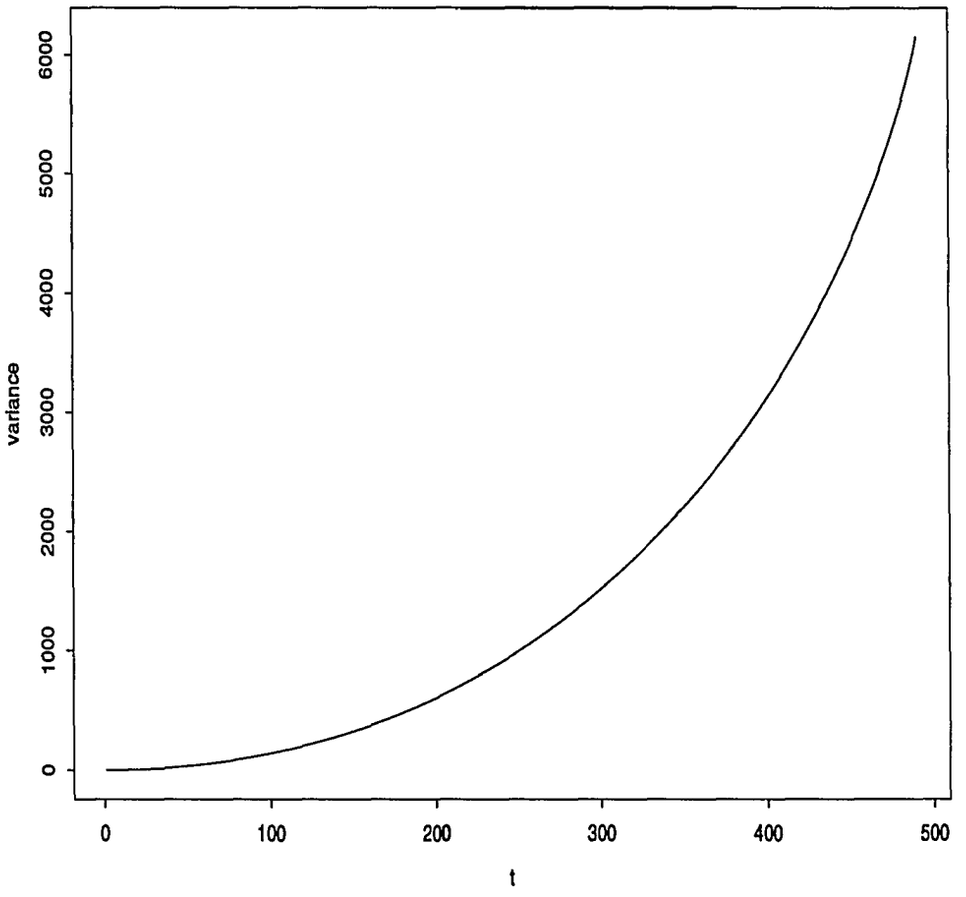

Figure 2.1: Variance of temporal K-function estimator, given by (2.11). Units of variance are t^2 and units of t are seconds.

In solving for the transformation, use initially only the first term in the Taylor series; i.e.,

$$\begin{aligned} g(y) &= c\int_0^y \left(\frac{t}{T\lambda^2}\right)^{-1/2} dt \\ &= c\left(\frac{-2}{\left(\frac{1}{T\lambda^2}\right)^{1/2}}\right)(y)^{1/2}. \end{aligned}$$

This implies the use of $g(y) = (y)^{1/2}$ as a first-order variance stabilizing transformation. For the data set discussed in Sections 2.4 and 2.5 we found that although this transformation does not completely stabilize the variance, much improvement occurs.

If the first two terms of the Taylor series expansion are used, the following function is obtained using Gradshteyn and Ryzhik (1980, p. 81):

$$g(y) \propto \log\left(\frac{2}{\lambda^2 T}\left(\frac{(1+4\lambda T)}{T}y + \frac{(1+4\lambda T)^2}{T^2}y^2\right)^{1/2} + 2\frac{(1+4\lambda T)}{T}y + 1\right).$$

Again using the application in Sections 2.4 and 2.5, we found that this transformation did not improve the stabilization of the variance much over the square root, so we recommend inference based on

$$\left(\hat{K}(t)\right)^{1/2}; \qquad t > 0.$$

2.3.2 Improving K function estimators using Stein's approach

Stein (1993) uses the following theorem to improve an existing d-dimensional K-function estimator.

Theorem 1 (Stein, 1993). Suppose $X_1, ..., X_n$ are i.i.d. and uniformly distributed on

$A \subset R^d$. For a given function $\phi(\cdot,\cdot)$,

$$\mathrm{var}_n\left(\sum_{i\neq j}\phi(X_i,X_j) - \sum_{j=1}^{n}[g(X_j) - a^{-1}I_g]\right)$$

is minimized by

$$g_0(x,\phi) = 2(n-1)\mathrm{E}[\phi(X_1,X_2)|X_1 = x] \tag{2.13}$$

and the minimizing variance is

$$2n(n-1)\left(a^{-2}S_2 - 2a^{-3}S_1 + a^{-4}S^2\right), \tag{2.14}$$

where

$$\begin{aligned} S &\equiv \int_{A^2}\phi(x,y)dxdy, \\ S_1 &\equiv \int_A\left[\int_A\phi(x,y)dy\right]^2 dx, \\ S_2 &\equiv \int_{A^2}\phi(x,y)^2dxdy, \\ I_g &\equiv \int_A g(x)dx, \end{aligned}$$

and $a = \nu_d(A)\square$.

Note that S_1 and S_2 are defined, for the temporal case, in (2.10) and (2.11) respectively. A detailed proof of this theorem is given in Appendix B. The proof does not depend on d and the arguments are the same for the temporal case as for the d-dimensional case. We now use this result to give a temporal analogue to Stein's estimator.

To estimate $K(t)$ we need an appropriate $\phi(\cdot,\cdot)$ which should be symmetric so that the variance calculation in (2.14) holds. Stein (1993) suggests dividing an unbiased estimator for $\lambda^2 aK(t)$ by (an estimate for) $\lambda^2 a$. Note that for any stationary

isotropic process in R^d there is a non-decreasing function $R(\cdot)$ on $[0,\infty)$ such that

$$K(t) = \omega_d \int_0^t r^{d-1} R(r) dr, \tag{2.15}$$

where ω_d is the surface area of the d-dimensional unit sphere (Ohser, 1983). Ohser (1983, Lemma 2.1) also shows that in general, for d-dimensional space,

$$\mathrm{E}\left(\sum_{i\neq j} \phi(X_i, X_j)\right) = \lambda^2 \int_0^\infty \int_A \int_\Omega \phi(x, x+r\gamma) I(x+r\gamma\epsilon A) \bar{\omega}(\gamma) d\gamma dx R(r) dr, \tag{2.16}$$

where Ω is the set of unit vectors on R^d, and $\bar{\omega}(\gamma)$ is the d-dimensional uniform density on Ω. To ensure that $\lambda^2 a\hat{K}(\cdot)$ is an unbiased estimator for any stationary isotropic process, we need to find a function $\phi(\cdot,\cdot)$ that satisfies the constraint formed by combining (2.15) and (2.16) in the following manner:

$$\mathrm{E}\left(\sum_{i\neq j} \phi(X_i, X_j)\right) = \lambda^2 a K(t) = \lambda^2 a \omega_d \int_0^\infty I(r<t) r^{d-1} R(r) dr. \tag{2.17}$$

We now reduce the constraint (2.17) to the temporal case. For $d = 1$, eq. (2.15) reduces to $K(t) = \int_0^t R(r) dr$. This is, however, the one-dimensional case rather than the temporal case. Since $K(t)$ is an expected value of occurrences close to an event x, and since the process is stationary and isotropic, equal numbers of events are expected to occur behind and in front of the event in question, implying that the temporal K function should be equal to one half of the one-dimensional K function. Thus, for the temporal case, $K(t) = \frac{1}{2}\int_0^t R(r) dr$ and eq. (2.16) reduces to

$$\begin{aligned} \mathrm{E}\left(\sum_{i\neq j} \phi(X_i, X_j)\right) &= \lambda^2 \int_0^\infty \int_A \frac{1}{2} \phi(x, x+r) I(x+r\epsilon A) dx R(r) dr \\ &\quad + \lambda^2 \int_0^\infty \int_A \frac{1}{2} \phi(x, x-r) I(x-r\epsilon A) dx R(r) dr. \end{aligned} \tag{2.18}$$

In the temporal case $a \equiv T$, so from (2.18) and (2.17) the constraint that $\phi(\cdot,\cdot)$ must satisfy in the temporal case is

$$\frac{1}{T}\int_A [\phi(x, x+r)I(x+r\epsilon A) + \phi(x, x-r)I(x-r\epsilon A)]dx = I(r<t). \tag{2.19}$$

Finding the optimal function g as defined in (2.13) under the constraint given in (2.19) is difficult so we use an approximation that has an exact solution. Suppose that the $\phi(\cdot,\cdot)$ used in the estimator (2.17) is uniformly bounded as $A \equiv [0,T]$ grows and that $\phi(x,y) = 0$ for $|x-y| > M$, where M is a fixed constant that does not depend on A (we can take $M = t$); then S, S_1, and S_2 are all $O(T)$. The expression for the variance of Stein's estimator, given by eq. (2.14), can then be written as:

$$2n(n-1)[T^{-2}S_2 + O(T^{-2})],$$

which allows us to concentrate on minimizing S_2 subject to (2.19). The left hand side of the constraint in (2.19) can be rewritten as follows:

$$\begin{aligned}
&\frac{1}{2T}\int_A [\phi(x, x+r)I(x+r\epsilon A) + \phi(x, x-r)I(x-r\epsilon A)]dx \\
&= T\left(\frac{1}{2}\int_A [\phi(x, x+r)I(x+r\epsilon A) + \phi(x, x-r)I(x-r\epsilon A)]\frac{1}{T^2}dx\right) \\
&= \frac{T}{2}\int_A [\phi(x_1, x_1+r)I(x_1+r\epsilon A) f_{12}(x_1, x_2 \mid |x_1-x_2| = r) f(r) dx_1 \\
&\quad + \frac{T}{2}\int_A \phi(x_1, x_1-r)I(x_1-r\epsilon A)] f_{12}(x_1, x_2 \mid |x_1-x_2| = r) f(r) dx_1 \\
&= T\mathrm{E}(\phi(X_1, X_2) \mid |X_1 - X_2| = r) f(r),
\end{aligned} \tag{2.20}$$

where $f_{12}(x_1, x_2 \mid |x_1 - x_2| = r)$ is the conditional density of X_1 and X_2 given that they are a distance r apart and $f(r)$ is the marginal density for $|X_1 - X_2|$. Using (2.20), the constraint shown in eq. (2.19) becomes

$$\mathrm{E}[\phi(X_1, X_2) \mid |X_1 - X_2| = r] = \frac{I[r<t]}{2Tf(r)}. \tag{2.21}$$

Constraint (2.21) fixes $\mathrm{E}[\phi(X_1, X_2)]$ so it is equivalent to minimize $\mathrm{var}[\phi(X_1, X_2)]$ subject to the constraint (2.21). Note that

$$\begin{aligned}\mathrm{var}[\phi(X_1, X_2)] &= \mathrm{E}[\mathrm{var}(\phi(X_1, X_2) \mid |X_1 - X_2|)] \\ &\quad + \mathrm{var}[\mathrm{E}(\phi(X_1, X_2) \mid |X_1 - X_2|)]. \qquad (2.22)\end{aligned}$$

But if $\phi(\cdot, \cdot)$ satisfies (2.21) then

$$\mathrm{var}[\mathrm{E}(\phi(X_1, X_2) \mid |X_1 - X_2|)] = \mathrm{var}\left[\frac{I(|X_1 - X_2| < t)}{2Tf(|X_1 - X_2|)}\right]$$

is fixed. The first term in (2.22) is always nonnegative and so is minimized at 0. This term equals zero if and only if $\phi(X_1, X_2)$ is measurable with respect to $|X_1 - X_2|$. So the solution to the problem of finding the optimal $\phi(\cdot, \cdot)$ is to find a symmetric function that satisfies the following conditions:

1. $\phi(x, y)$ is uniformly bounded as T increases;
2. $\phi(x, y) = 0$ for $|x - y| > t$;
3. $\phi(x, y)$ is measurable with respect to $|x - y|$;
4. $\phi(x, y)$ satisfies (2.19).

By satisfying these four conditions, an (approximately) optimal solution is found.

2.3.3 Improving the initial estimator

We start with the following function based on the temporal analogue to Ohser and Stoyan's estimator for the K function (1981), discussed in Section 2.3.1. Define

$$\phi(X_1, X_2) = \frac{TI(|X_2 - X_1| < t)}{2(T - |X_1 - X_2|)}. \qquad (2.23)$$

See Appendix C for a proof that this function satisfies the conditions laid out in Section 2.3.2.

The first step in constructing the new estimator of the K function is to find the function $g_0(x,\phi)$ referred to in eq. (2.13) of Theorem 1. Recall that X_1 and X_2 are independent and uniformly distributed on the region of interest $A \equiv [0,T]$. Since X_1, X_2 are independent we have that

$$f(x_2|X_1 = x) = \frac{1}{T} I(0 \le x_2 \le T),$$

from which

$$\begin{aligned} \mathrm{E}[\phi(X_1,X_2)|X_1 = x] &= \frac{1}{T}\int_0^T \phi(x,x_2)dx_2 \\ &= \int_0^T \frac{I(|x_2 - x| < t)}{2(T - |x - x_2|)}dx_2 \end{aligned}$$

$$= \begin{cases} \int_0^x \frac{1}{2(T-x+x_2)}dx_2 + \int_x^{x+t} \frac{1}{2(T+x-x_2)}dx_2 & \text{, for } 0 < x < t \\ \int_{x-t}^x \frac{1}{2(T-x+x_2)}dx_2 + \int_x^{x+t} \frac{1}{2(T+x-x_2)}dx_2 & \text{, for } t < x < T-t \\ \int_{x-t}^x \frac{1}{2(T-x+x_2)}dx_2 + \int_x^T \frac{1}{2(T+x-x_2)}dx_2 & \text{, for } T-t < x < T. \end{cases}$$

Therefore,

$$g_0(x,\phi) = \begin{cases} (n-1)[2\log(T) - \log(T-t) - \log(T-x)] & \text{, for } 0 < x < t \\ (n-1)[2\log(T) - 2\log(T-t)] & \text{, for } t < x < T-t \\ (n-1)[2\log(T) - \log(T-t) - \log(x)] & \text{, for } T-t < x < T. \end{cases} \quad (2.24)$$

The final piece needed is

$$\begin{aligned} I_{g_0} &= \int_A g_0(x)dx \\ &= \int_0^t (n-1)[2\log(T) - \log(T-t) - \log(T-x)]dx \end{aligned}$$

$$
\begin{aligned}
&+ \int_t^{T-t} (n-1)[2\log(T) - 2\log(T-t)]dx \\
&+ \int_{T-t}^{T} (n-1)[2\log(T) - \log(T-t) - \log(x)]dx \\
= \; & (n-1)[2t\log(T) - t\log(T-t) + (T-t)\log(T-t) - (T-t) - T\log(T) \\
&+ T] + (n-1)[2T\log(T) - 4t\log(T) - 2T\log(T-t) + 4t\log(T-t)] \\
&+ (n-1)[2t\log(T) - t\log(T-t) - T\log(T) + T + (T-t)\log(T-t) \\
&- (T-t)] \\
= \; & 2(n-1)t. \qquad (2.25)
\end{aligned}
$$

Theorem 1 suggests as the estimator of the K function:

$$\hat{K}(t) = \frac{1}{\lambda^2 T}\left[\sum_{t_i \neq t_j} \phi(t_i, t_j) - \sum_{t_i}[g_0(t_i) - T^{-1} I_{g_0}]\right],$$

which is unbiased for all stationary point processes. However, in most situations, λ is unknown and must be estimated by $\lambda = (n/T)$. Thus, we propose

$$\hat{K}(t) = \frac{T}{n^2}\left[\sum_{t_i \neq t_j} \phi(t_i, t_j) - \sum_{t_i}[g_0(t_i) - T^{-1} I_{g_0}]\right].$$

Stein(1993) shows from simulations that the bias of this estimator is much smaller for the homogeneous Poisson process than for the Poisson cluster process.

Using the $\phi(\cdot,\cdot)$ defined in (2.23) and the results (2.24) and (2.25) we have

$$
\begin{aligned}
\hat{K}(t) \; = \; & \frac{T}{n^2}\left\{\sum_{t_i \neq t_j} \frac{TI(|t_i - t_j| < t)}{2(T - |t_i - t_j|)}\right. \\
& \left. - (n-1)\left[\sum_{0<t_i<t} [2\log(T) - \log(T-t) - \log(T-t_i)]\right.\right.
\end{aligned}
$$

$$+ \sum_{t \le t_i \le T-t} [2\log(T) - 2\log(T-t)]$$

$$+ \sum_{T-t < t_i \le T} [2\log(T) - \log(T-t) - \log(t_i)] - n\frac{2t}{T}\Bigg]\Bigg\}.$$

This can be simplified somewhat to:

$$\begin{aligned}\hat{K}(t) &= \frac{T}{n^2}\Bigg\{\sum_{t_i \neq t_j} \frac{TI(|t_i - t_j| \le t)}{2(T - |t_i - t_j|)} \\ &\quad - (n-1)\Bigg[2n\log(T) - n\log(T-t) - \sum_{0<t_i<t} \log(T-t_i) \\ &\quad - \sum_{t \le t_i \le T-t} \log(T-t) - \sum_{T-t<t_i\le T} \log(t_i) - n\frac{2t}{T}\Bigg]\Bigg\}. \qquad (2.26)\end{aligned}$$

In Appendix D we verify that, conditional on n,

$$\mathrm{var}_n(\hat{K}(t)) = \frac{2T^2(n-1)}{n^3}\left[\left(\frac{1}{2} - \frac{2(T-t)}{T}\right)\log\left(\frac{T}{T-t}\right) + \frac{2tT + t^2}{T^2}\right]. \qquad (2.27)$$

This variance is increasing in t as can be readily seen in Figure 2.2.

To find a transformation of the K-function estimator that has (at least relatively) constant variance under CTR, we again apply the δ-method. Let $V(t)$ be the variance function given in (2.27). Then

$$V(t) \approx \frac{5T(n-1)}{n^3}t + \frac{5(n-1)}{n^3}t^2.$$

Solving the following integral for $g(y)$ should give an approximate variance-stabilizing transformation:

$$g(y) = \int_0^y \frac{c}{(V(x))^{1/2}}dx$$

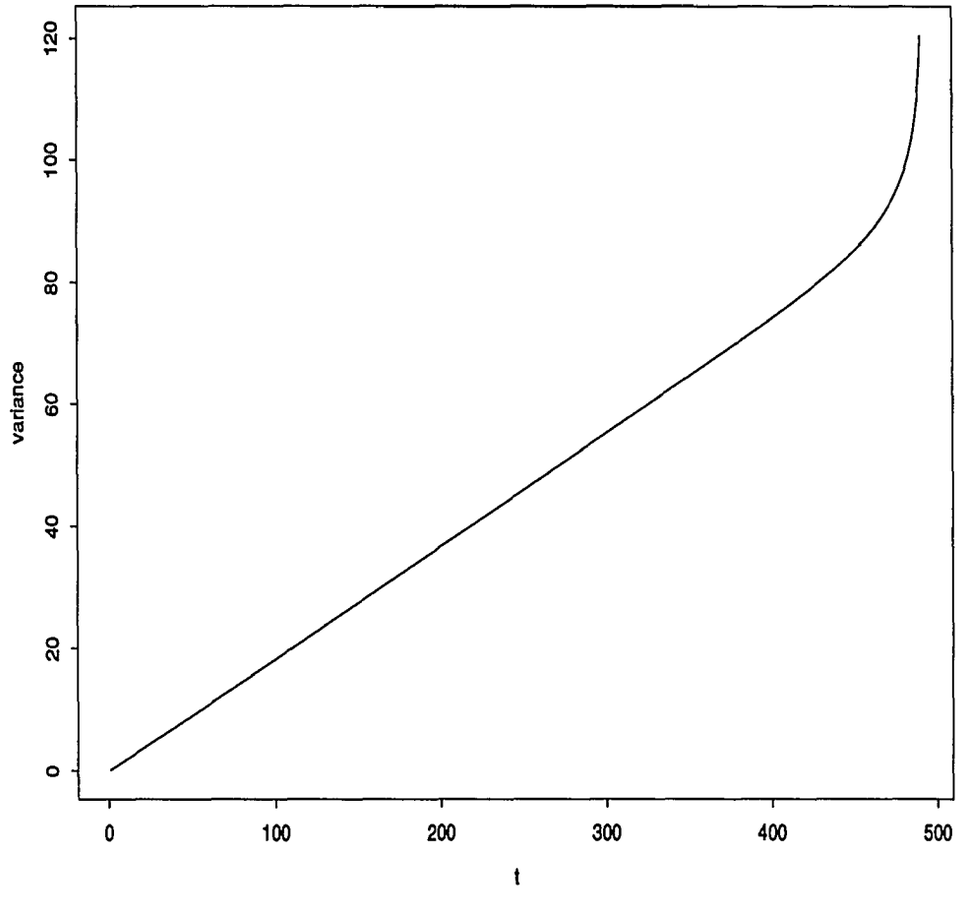

Figure 2.2: Variance of improved temporal K-function estimator, given by (2.26). The variance is in units of t^2 and units of t are seconds.

Using only the first term of the Taylor series this gives:

$$g(y) = 2c\left(\frac{n^3}{5T(n-1)}y\right)^{1/2},$$

from which the transformation

$$g(y) = (y)^{1/2}$$

is suggested. The transformation becomes quite a bit more complicated if the first two terms of the Taylor series are used, giving

$$\begin{aligned} g(y) &= \left(\frac{n^3}{5(n-1)}\right)^{1/2} \log\left(2\left(\frac{25T(n-1)^2}{n^6}y + \frac{25(n-1)^2}{n^6}y^2\right)^{1/2}\right. \\ &\quad \left. + \frac{10(n-1)}{n^3}y + \frac{5T(n-1)}{n^3}\right), \end{aligned}$$

which suggests the transformation:

$$g(y) = \log\left(2\left(y(T+y)\right)^{1/2} + 2y + T\right).$$

Again, we will use the simpler square root transformation. The variance for $\hat{K}(t)^{1/2}$, while not constant, increases at a much slower rate than the variance for $\hat{K}(t)$, and is represented approximately by

$$\text{var}\left(\hat{K}(t)^{1/2}\right) \approx \left[\frac{\partial}{\partial y}g(y)\Big|_{y=(\mathrm{E}(\hat{K}(t)))}\right]^2 V(t),$$

where $V(t)$ is given by eq. (2.27). Therefore,

$$\begin{aligned} \text{var}\left(\hat{K}(t)^{1/2}\right) &\approx \left(\frac{1}{2t^{1/2}}\right)^2 V(t) \\ &= \frac{T^2(n-1)}{2tn^3}\left[\left(\frac{1}{2} - \frac{2(T-t)}{T}\right)\log\left(\frac{T}{T-t}\right) + \frac{2tT+t^2}{T^2}\right] \quad (2.28) \end{aligned}$$

We illustrate in Figure 2.3 that the (approximate) variance for $\hat{K}(t)^{1/2}$, as given by (2.27), is indeed much more stable than that of $\hat{K}(t)$; cf Figure 2.2.

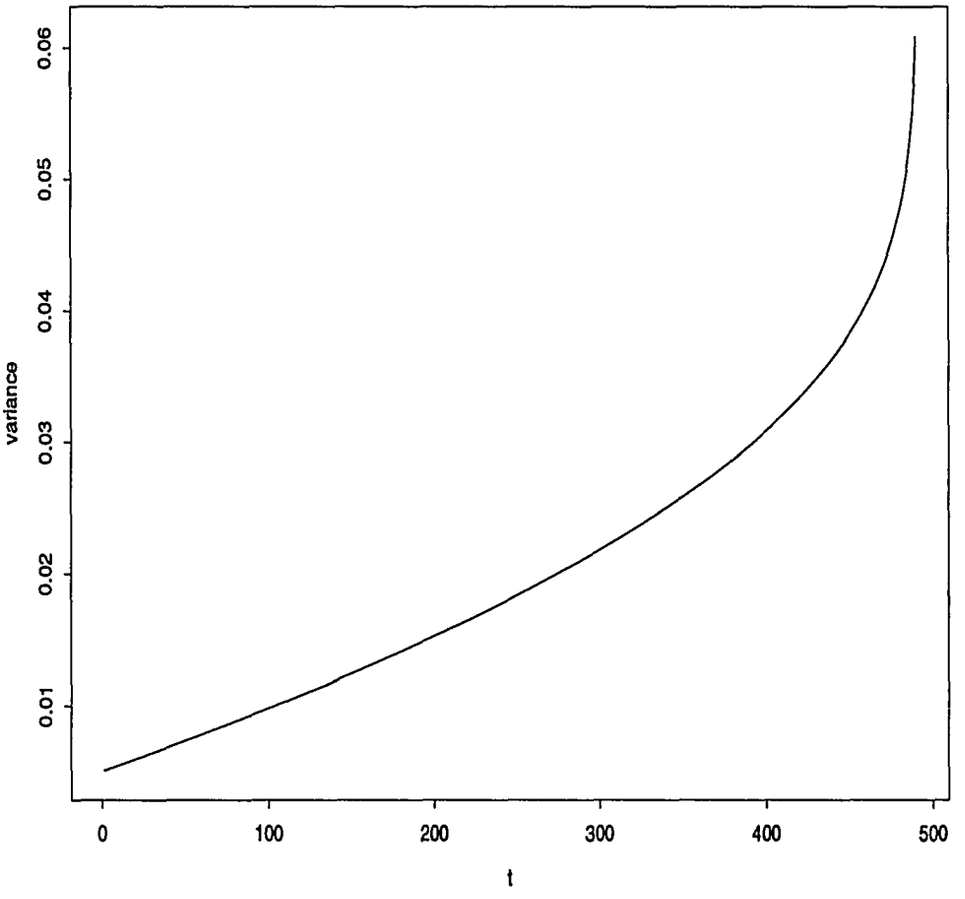

Figure 2.3: Variance of the transformed temporal K-function estimator, $\hat{K}(t)^{1/2}$. Units of the variance are t, units of t are seconds.

2.4 An Example of the Use of Temporal K Functions

2.4.1 Preliminary explanation of the data set

The following application is an example where temporal K functions offer different, and possibly more important information than traditional statistical approaches. Stated simply, the problem is to find possible physiological/neurological side effects of drugs (e.g., Mullenix, Kernan, Tassinari and Schunior (1989)), but in the early stages of testing it is unknown what type of side effect might occur in any particular experiment. This has led to an exploratory stage where the researchers look for behavioral changes, believing that such changes point to the occurrence of some type of physiological/neurological change.

Experiments were run by separating rats into two groups, a control group and an exposed group, and at some relevant point in time observing the rats in pairs, one control and one exposed, for 15 minutes in a novel environment. The pairing is not done because of similarities between rats, but rather, to help control for environmental differences across observation times. The actions for each rat, for each second of the time period are recorded by a computer pattern recognition program leading to an observational vector of length 900, each unit of which represents the value of a categorical variable, "body position," corresponding to the rat's act at the beginning of that second. The possible values for this variable are: stand, sit, walk, rear, and lying down. Originally, researchers looked at statistics representing the total time the rat spent performing each act (or body position). They then added a statistic that represented the number of times that each act was initiated.

Wender (1971) showed that some types of behaviors, such as hyperactivity in

children, will not be discovered using only these measures. Hyperactivity has been traditionally seen as a state of increased motor activity, but this sense of increase comes from the behavior occurring in a different pattern than normal. The temporal K function is useful in helping to identify differences in the patterns of events. In this article we demonstrate the use of the temporal K function by analyzing the act "stand." We start, in Section 2.4.2 by analyzing the K-function estimates of just one group of rats and we show that the null hypothesis of CTR is an appropriate model for a control group. We go on in Section 2.5 to give a technique for using the K function to detect differences between two groups of rats.

For this analysis, the events of interest are the *initiations* of the act stand. The duration of time that the rat stood continuously is not of interest here. Thus we concentrate on patterns of stand initiations. To do this, the original data vector was first recoded with 0's and 1's, where 1 represented the act stand, 0 anything else. We then eliminated excess "stand times" by deleting all 1's occurring consecutively just after the initial 1. The time periods deleted could not possibly contain an initiation since, to initiate an act, the rat must first have been doing something else immediately preceding the initiation. The final data vector is a string of 0's and 1's with the 1's occurring as singletons. Using this technique, the original data vector can be transformed to allow analysis for any one particular act.

2.4.2 Analysis

Each rat has its own estimated K function which can be calculated using eq. (2.26). As pointed out in Section 2.3.3, this estimator has a rapidly increasing variance as t increases. Instead of using $\hat{K}(t)$, it is better to use the square-root transfor-

mation. For this application, the null hypothesis is lack of pattern in the initiations, or CTR. Under CTR, we have shown in Section 2.3 that $K(t) = t$, which leads to $[K(t)]^{1/2} = t^{1/2}$. It is common, instead of using the K function or its square root, to use the function

$$L(t) \equiv [K(t)]^{1/2} - t^{1/2}. \tag{2.29}$$

Under CTR, $L(t) = 0$ for all t, which means that departures from CTR are easily interpretable and seen graphically. Using the L-function estimate based on eq. (2.29) and eq. (2.26), departures from CTR are often consistently positive or consistently negative: A positive departure implies that the events are more clustered, while a negative departure implies that the events occur in a more regular fashion. In reality, the L-function estimate will have a small bias because the K-function estimate has a small bias (Stein, 1993) and because of the nonlinear square root transformation, but this bias should be minor.

In preparation for testing the null hypothesis of CTR for control rats, K-function estimates were calculated using eq. (2.26) for 20 rats that were observed as controls with no treatment whatsoever. These K-function estimates were then transformed into L-function estimates using eq. (2.29). Under CTR, eq. (2.28) gives the variance of the L-function estimator. However, λ must be estimated, here by n/T, and since each rat had a different total time after the original data vector was recoded, each L-function estimate will have a different estimated variance. For comparison purposes, the 20 L-function estimates were standardized by dividing each estimate by its standard deviation at every time point; the standard deviation was computed using eq. (2.28).

Before proceeding with the more formal statistical analysis, we provide a graphi-

cal analysis. Figure 2.4 shows the 20 standardized L-function estimates for this group of control rats, as well as the theoretical value (namely zero) for the L function under CTR. Several points need to be made about this plot. First, the L-function estimates appear to be close to the line at zero but, there are several L functions that are quite positive. These are indicative of a possibly troublesome component of error caused by rat-to-rat variability.

We now address the question of how to judge analytically whether CTR is an appropriate model for the control rats. We use a signed area measure between each estimate and the hypothesized L function at 0. This area is computed as follows: the area where the estimate lies above 0 is positive, the area where the estimate is below 0 is negative, and the two areas are then summed. Under CTR, the L-function estimates should hover around 0, and, therefore, the average signed area should be close to 0.

It was hoped that the distribution of the average of the 20 areas would be something close to normal so that the normal distribution could be used in the test of the null hypothesis that the average area is zero. To judge the normality of the distribution of the average signed area, 5000 bootstrap samples of size 20 were chosen with replacement from the original sample, and the 5000 average signed areas were calculated. From the histogram in Figure 2.5(a), the distribution appears to be slightly right skewed but the normal plot in Figure 2.5(b) confirms that the distribution of the averages is approximately normal.

To test whether the pattern of stand initiations in control rats follows CTR, a standard one-sample hypothesis test was performed on the set of 20 areas, where the standard deviation of the bootstrapped means was used as an estimate of the

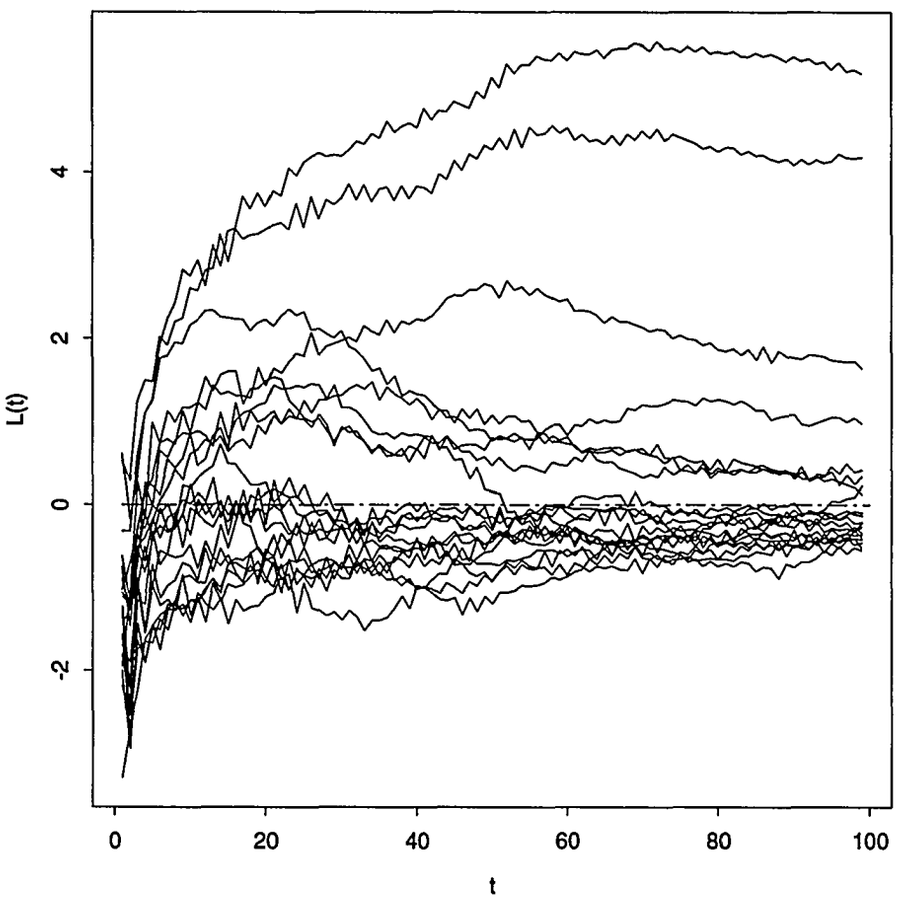

Figure 2.4: Standardized L-function estimates for 20 control rats. Each solid line represents one rat. The dashed horizontal line is the L function under CTR. Units on the horizontal axis are in seconds and units on the vertical axis are in $(\text{seconds})^{1/2}$.

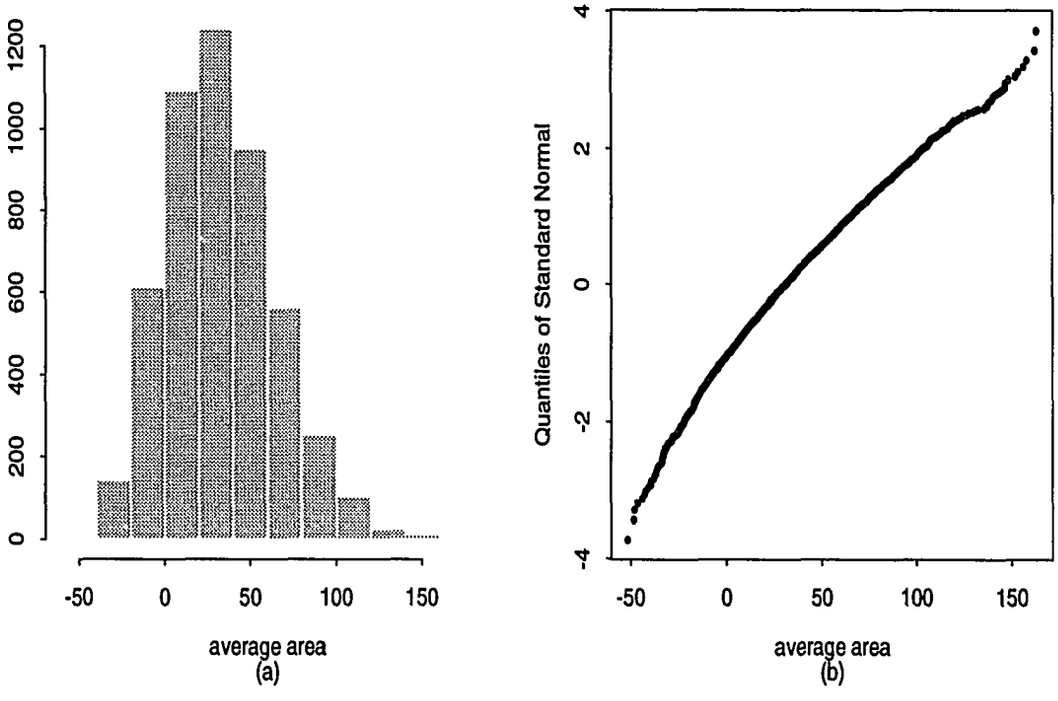

Figure 2.5: Two figures for a sample of 20 control rats. The histogram in (a) shows the bootstrapped distribution of the average area between the standardized L-function estimates and CTR, while the normal probability plot in (b) compares this distribution to the normal distribution.

true standard deviation of the average area. The usual t-statistic with the sample standard deviation in the denominator is not acceptable in this case because of the occasional rat with an unusual L function. In the presence of outliers, the t-statistic will not have the usual t-distribution, but often, a skewed distribution, and therefore cannot be used as easily. By using the standard deviation based on the bootstrapped sample, we get a much more stable estimate for the true parameter than the sample standard deviation, allowing the z-score to be approximately normally distributed (Hall, 1988). The z-score for the test was 1.018 with a p-value of .3087. The results suggest that the mean of the average signed area between the L-function estimates and CTR's L function is zero and, therefore, that a CTR model is appropriate.

2.5 Comparing Two Sets of K Functions

In Section 2.4, we gave a method for comparing a sample of L-function estimates to an hypothesized value. We also established that CTR seems to be a reasonable assumption when describing the events of stand initiations in a group of control rats. However, most often the researchers wish to judge the effects of a specific treatment on the behavior of the rats. Therefore, an exposed (to treatment) group and control group of rats are compared; the rats in the control group have, themselves, had some sort of treatment (i.e. injected with saline solution when the "treated" rats were injected with the drug of interest). The result is that the control rats may not be exhibiting truly normal behavior. If normal behavior and CTR are equated, we might expect to see some departure from CTR, even among control groups. So, for this type of experiment, a method is needed to compare estimates from the control group with those from the exposed group. We now provide a method that can be

used when the data are paired; that is, one control observation is paired with one treated observation.

There are two possible approaches to comparing two sets of (standardized) L-function estimates. The first would be to compute some sort of area between each pair of estimates. The average signed area would be zero under the null hypothesis of no difference, so a test based on the average signed area would be easy to use and interpret. However, there are cases where this approach would not pick up potentially interesting differences. For example, one of the pair of rats could have a very stable L function while the other rat's L function might oscillate above and below the stable estimate. A test based on the average absolute area is difficult to implement, because it is unclear what the mean value is under the null hypothesis of no difference.

The second approach (and the one we take) is to compute areas between each of the rats' standardized L-function estimate and a third entity (for this example, the CTR L function), compute the difference in area for each of the pairs and average these differences across pairs. In an attempt to pick up any type of difference, we use two types of areas separately, both the positive (the area where the estimate lies above the CTR L function) and the negative (the area where the estimate lies below the CTR L function). We chose these areas rather than the signed and absolute areas because of their interpretation. Generally, L functions that are positive represent events that are more clustered, while L functions that are negative represent events that are more regular. It should be noted that the signed and absolute areas are simply linear functions of the positive and negative areas. The method results in two test statistics, the average difference between positive areas and the average difference between negative areas. The test of no difference between control and treatment rats

is then based on two-sided paired comparisons of the two types of areas.

We shall use two different data sets in this section, a control-control data set, and an exposed-control data set. First, the data used in Section 2.4 is actually one half of a control-control data set; recall that a z-score of 1.018 and a p-value of .3087 were obtained. The hypothesis test from Section 2.4 was also applied to the second half of the data set with the resulting z-score of .2711 and p-value of .7863. This indicates that both groups of the control-control can be modeled by CTR, as expected. The second data set comes from an experiment where each of the 20 pairs of rats consisted of one rat injected with 1 mg/kg of d-amphetamine, a drug known to cause behavioral differences, and the other rat was injected with an equal amount of saline solution. The hope is to detect a difference due to the d-amphetamine while controlling for the effects of the injection.

Figure 2.6 gives an idea of how much difference can be expected in the L-function estimates from a control-control pair as opposed to an exposed-control pair. Figure 2.6(a) is a plot of the standardized L-function estimates of a typical pair of rats in a control-control experiment, whereas Figure 2.6(b) shows a typical pair of rats in an exposed-control experiment. It is clear that while some difference can be expected between two control rats, this difference is not nearly as large as the difference found between an exposed and a control rat. Figures 2.7(a) and (b) show the two samples separately. These two figures indicate that the exposed rats have events that tend to be more clustered than the control rats.

We now illustrate the proposed method of analysis. As in Section 2.4, we used a bootstrap approach to find the distribution of the two average differences in areas. We did this for both statistics, in both data sets, using 5000 bootstrap samples. We start

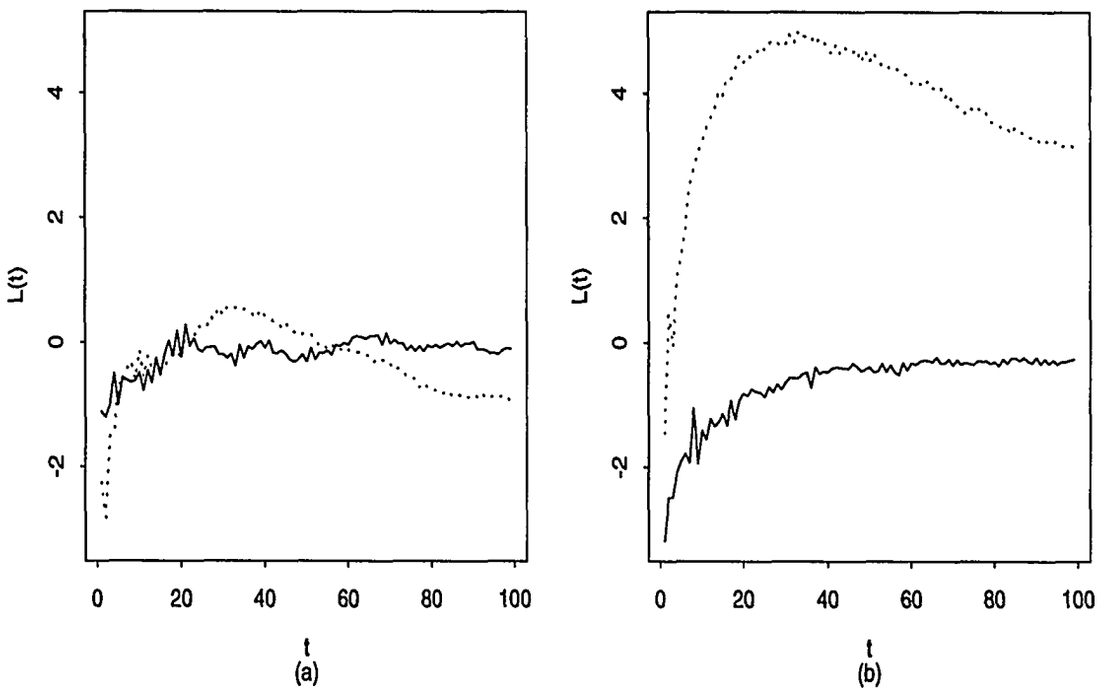

Figure 2.6: Two pairs of standardized L-function estimates from rat behavior data. Plot (a) shows the L-function estimates from a typical pair of rats in a control-control experiment. Plot (b) shows the L-function estimates from a typical pair of rats from an exposed-control experiment. Units on the axes are the same as for Figure 2.4.

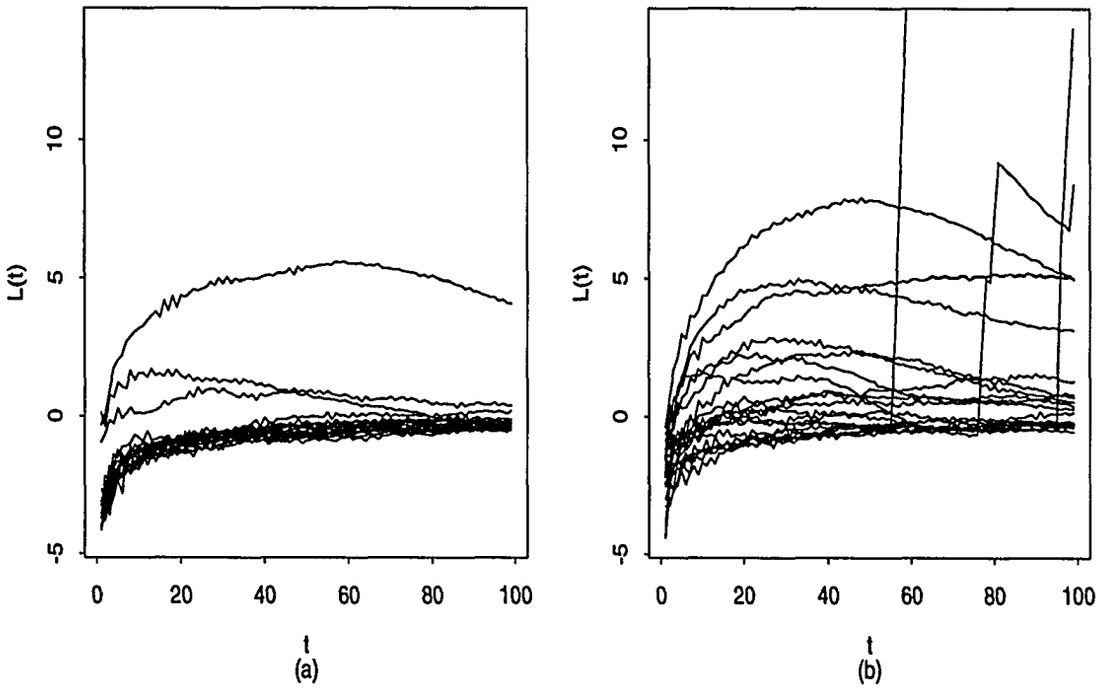

Figure 2.7: Two plots of standardized L-function estimates. Plot (a) shows the estimates for 20 control rats; plot (b) shows the estimates for 20 exposed rats. Units on the axes are the same as for Figure 2.4.

by examining the control-control results. Figures 2.8(a) and (b) give the histograms for the average difference in the positive area and average difference in the negative area, respectively, while Figures 2.8(c) and (d) give the normal probability plots for these same distributions. All four plots indicate that the two average differences are, at least approximately, normally distributed. This means that the usual two-sided, paired-comparison hypothesis test using the normal distribution is acceptable; the standard deviation of the 5000 bootstrapped values is used as the estimate of the true standard deviation.

Figures 2.9(a), (b), (c), and (d) give the same plots as those in Figure 2.8 but for the bootstrapped distribution based on the exposed-control data set. Although the average difference in positive area shows a little skewness to the right, the normal plot is not far from a straight line, again indicating that the tests described above can be used.

For the two data sets under consideration, the tests gave the following results: The control-control experiment has z-score for the mean positive area given by $z = .3906$ with a p-value of .6961, and for the mean negative area $z = .9376$ with a p-value of .3485. Some care must be taken because of multiple testing on a single data set, so we use Bonferroni's inequality to provide a conservative testing procedure. To test the null hypothesis of no difference at $\alpha = .05$, we compare the p-values to .025. Clearly, in the case of the control-control data set, there are no detectable differences.

The exposed-control set has z-score for the mean positive area given by $z = 2.012$ with $p = .0442$, and for the mean negative area $z = -3.668$ with $p = .0002$. Again we use the Bonferroni inequality and compare the p-values to .025, in order to provide an overall significance level of $\alpha = .05$. For this set of data, we find a statistical

difference in the average difference in negative area. The differences were computed as *exposed area−control area* so with a negative z-score for the difference in negative area, the control rats had more negative area than the exposed rats. In other words, the exposed rats had events of stand initiations that were more clustered. We also note that the p-value for the difference in positive area, though not statistically significant at $\alpha = .05$, is not far from being significant. The Bonferroni inequality guarantees that significance in any one of the tests results in an overall significance, and hence we declare the exposed rats' stand initiations to be significantly more clustered than those of the control rats.

These two data sets illustrate that the tests based on differences of L functions (and subsequent bootstrapping), have low error rates. Not only are we able to detect differences between L-function estimates of exposed and control groups of subjects, we are also able to interpret the types of departures observed.

2.6 Discussion

In this article, we have adapted the K function (and L function) to a one-dimensional setting, in which there already exist many classical statistical methods for the analysis of point processes (e.g., Cox and Lewis, 1966; Snyder, 1975; and Karr, 1986). We now discuss the advantages of our approach over the more classical methods. Most are based on analyzing the length of times between points in the process (interarrival times), usually by fitting them to a distributional model or testing their fit to such a model. The usual model fitted is a Poisson process, which results in an exponential model for the interarrival times. The problem with this approach is that it does not take into consideration the patterns at different scales;

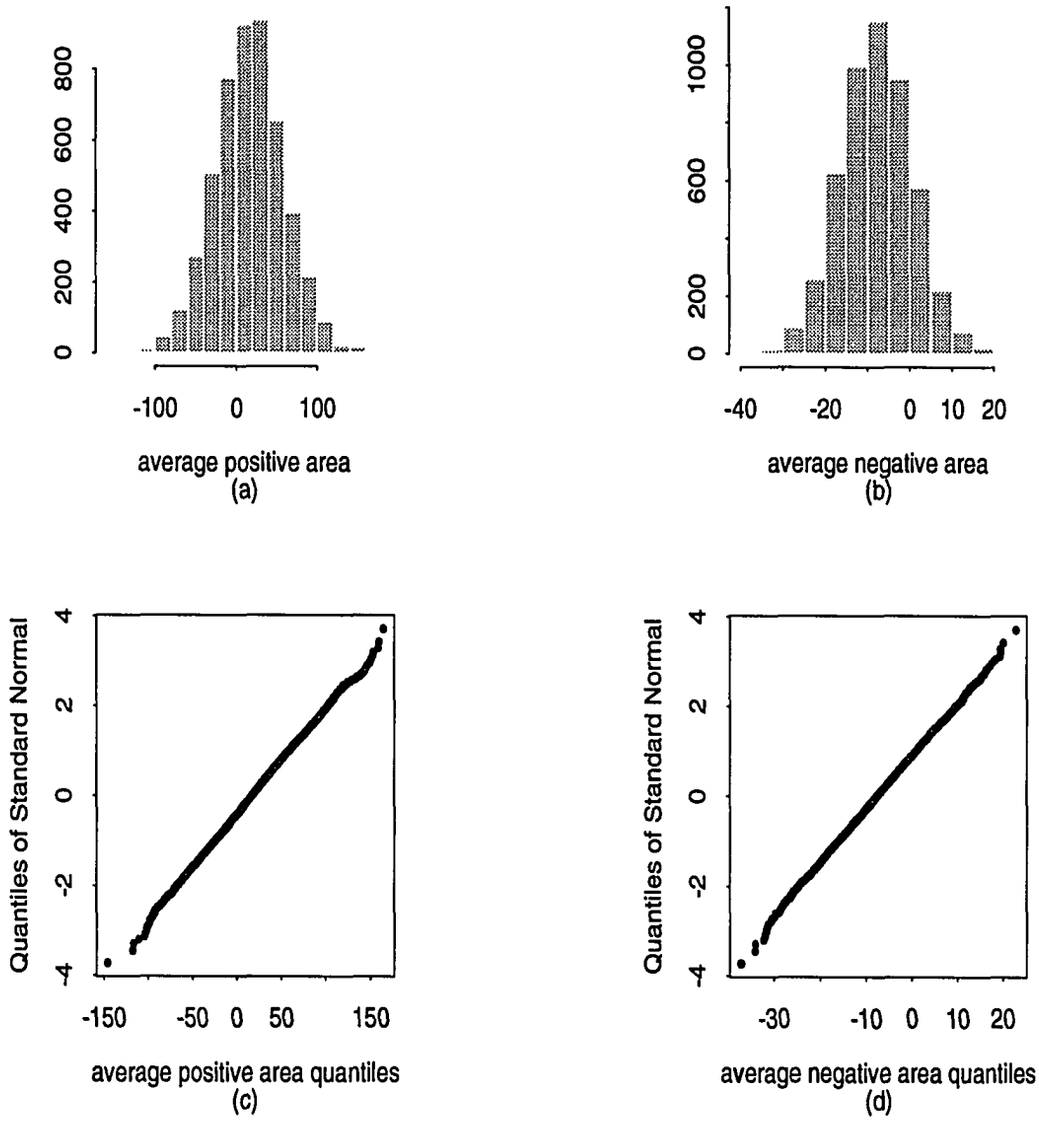

Figure 2.8: Four figures comparing 20 pairs of control rats. The histogram in (a) shows the bootstrapped distribution of the average difference in positive area between the standardized L-function estimates and CTR's L function, and the histogram in (b) shows the bootstrapped distribution of the average difference in negative area between the standardized L-function estimates and CTR's L function. Plots (c) and (d) give the corresponding normal probability plots.

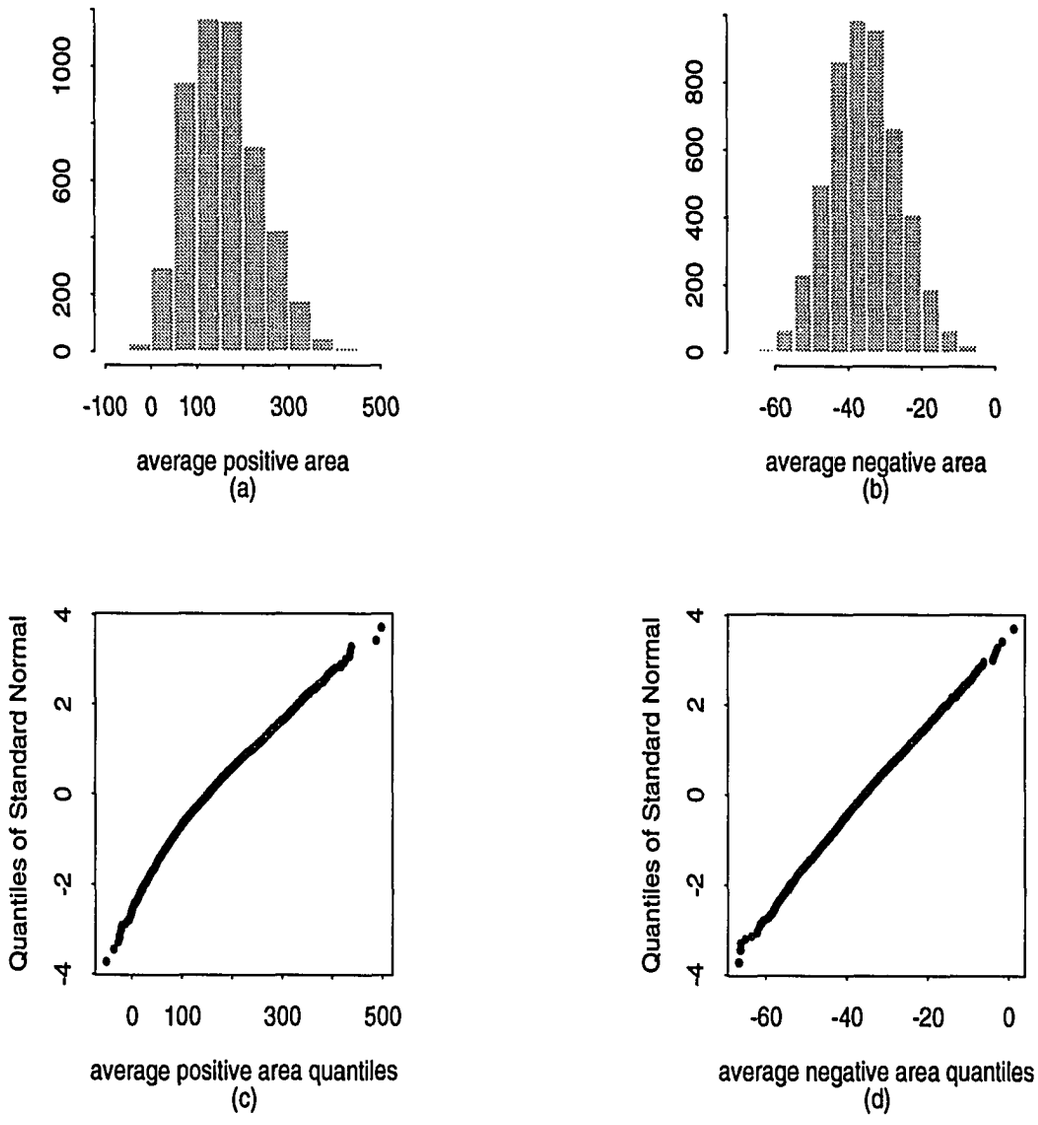

Figure 2.9: Four figures comparing 20 pairs of rats, one in each pair is treated, the other control. The histogram in (a) shows the bootstrapped distribution of the average difference in positive area between the standardized L-function estimates and CTR's L function, and the histogram in (b) shows the bootstrapped distribution of the average difference in negative area between the standardized L-function estimates and CTR's L function. Plots (c) and (d) give the corresponding normal probability plots.

its concern is only with the small-scale properties of the point process and it has difficulty detecting structure of a global nature. On the other hand, the K function, $K(t)$; $t \geq 0$, is able to distinguish between patterns of the process at many different scales t.

There is one statistic offered by Cox and Lewis (1966) that does, at least, alleviate the problem of only looking at the small scale. If $N(t) \equiv$ number of points in $[0, t]$, then the variance-time curve is defined as $V(t) \equiv \text{var}(N(t))$; $t \geq 0$. In fact, from Section 4.5 of Cox and Lewis (1966), we obtain

$$\begin{aligned} V(t) &= \lambda t - \lambda^2 t^2 + 2\int_0^t \int_0^v \lambda \Pr(\text{event in } (u, u+du] \mid \text{event at } 0) dv \\ &= \lambda t - \lambda^2 t^2 + 2\lambda^2 \int_0^t K(v) dv. \end{aligned}$$

Hence

$$K(t) = -\frac{1}{2\lambda} + t + \frac{1}{2\lambda^2} V'(t)$$

and so, in principle, knowledge of the variance-time curve is equivalent to knowledge of the K function. Because much is known about estimation of $K(t)$, as well as hypothesis testing based on these estimators, the need for obtaining analogous results for $V(t)$ is not so impelling.

In general, the classical methods presented by Cox and Lewis (1966) have tests associated with them which assume the data come from one realization of a point process and a hypothesized theoretical distribution of this point process. Some of the tests, notably the Kolmogorov-Smirnoff test, based on the interarrival times have been adapted somewhat to test the null hypothesis that two realizations of a point process follow the same distribution, by using the empirical cumulative distribution function (CDF) for the second process in place of the hypothesized CDF (Lindgren,

1976, p. 494). Karr (1986) also discusses estimation of a point process based on several sample realizations of that point process. Nevertheless, it would appear that none of the classical methods is currently able to do all that we have developed for the K function in Sections 2.4 and 2.5.

In Section 2.4 we developed a method of analysis using L-function estimates to compare one group of subjects to the null hypothesis of complete temporal randomness (CTR). We computed the signed area between each L-function estimate and the L function of CTR. We then tested the null hypothesis that the true average area is zero. We concluded that the normal behavior of rats, as measured by the pattern of the initiations of the act "stand," can be modeled adequately by CTR.

In Section 2.5, we gave a method of analysis for comparing two groups of subjects to each other. This analysis consisted of computing both the positive and the negative area between each subject and the L function of CTR. Then, the difference in area (for each type of area) between each pair of subjects was computed and hypothesis tests were performed to test the null hypothesis no average difference in area. We showed that, in a control-control experiment, this method finds no difference and in an exposed-control experiment where a difference is expected, one is found.

It should be noted that the data sets we were using have samples of size 20. In looking at the average difference in areas, the central limit theorem is probably having an effect. We have no reason to believe that the differences are themselves normally distributed. In fact, as noted in Section 2.4, we expect some outliers indicating non-normal data. So, before using these methods on smaller samples, some diagnostics should be performed. At the least, bootstrapping, as described in Section 2.4, should be done to assess the possibility of normality for the average difference. For the

case of only one subject, confidence bounds around the L-function estimate can be constructed using the variance function given by eq. (2.28).

The examples used in Sections 2.4 and 2.5 illustrate a situation where researchers have known that traditional forms of analysis do not capture differences very well. It has been shown (Wender, 1971) that some types of behavior are characterized by a change in the pattern of act initiations. A K-function analysis will pick up such differences between two groups of subjects (or between one group of subjects and complete temporal randomness). In fact, researchers in this field began doing a type of analysis using K functions several years. Cressie (1991) discusses a similar experimental situation but using monkeys and gives a temporal analogue to a different K-function estimator. Kernan, Mullenix, Kent, Hopper, and Cressie (1988) go one step further. These researchers used a temporal analogue of an estimator of the K function from Diggle (1983) and tested for differences between two groups of rats. First they found an average K-function estimate for each of the two groups. Then, to judge if those two averages were different, they picked eight values of t along the positive axis, calculated a test statistic at each of the eight points, and finally judged the two averages to be different if three or more points in a row showed a difference in the same direction. Kernan and Meeker (1992) used the same type of analysis in combination with several other types of more traditional analyses to get a broader view of the data and a more definitive answer to the question of whether any type of difference exists between the behaviors of the two groups of rats.

We feel that we have improved on the analysis using the K function found in Kernan et al. (1988) and Kernan and Meeker (1992) in two ways. First, we have found an estimator that is not only better suited to the temporal case, but also an

approximate best estimator. Second, we have improved on the method of analysis by finding a way to incorporate all 40 estimates and for all t, in our testing procedure.

Finally, we note several areas for further research. Kernan et al. (1988) used not only the K function but also a form of the cross K function as a way of looking for independence between the patterns of two acts. The derivation of an approximate best estimator for the cross K function would most likely follow similarly to that presented in Section 2.3, and likewise for the method of analysis presented in Sections 2.4 and 2.5. We leave the verification to future work. We also note that, while the experiment described in this article is well-suited for this type of analysis, potentially, any experiment with an event process occurring over time could benefit from such a K-function analysis.

2.7 Acknowledgments

The authors would like to thank a referee for suggestions that led to an improved discussion in Section 2.6. This article forms part of the first author's Ph. D. Dissertation in the Department of Statistics, Iowa State University. The research was partially supported by the Office of Naval Research (N00014-93-1-0001), the National Science Foundation (DMS-9204521), and the National Security Agency (MDA904-92-H-3021).

2.8 Bibliography

Bishop, Y.M.M, Feinberg, S. E., and Holland, P.W. (1975) *Discrete Multivariate Analysis.* Cambridge, Massachusetts: The MIT Press.

Cox, D.R., and Lewis, P.A.W. (1968) *The Statistical Analysis of Series of Events.* London: Methuen & Co., Ltd.

Cressie, N. A. C. (1991) *Statistics for Spatial Data.* New York: Wiley - Interscience.

Daley, D.J., and Vere-Jones, D. (1972) A Summary of the Theory of Point Processes. In *Stochastic Point Processes* (P.A.W. Lewis, ed.). New York: Wiley, pp. 299-383.

Diggle, P.J. (1983) *Statistical Analysis of Spatial Point Patterns.* New York: Academic Press.

Gradshteyn, I. S., and Ryzhik, I. M. (1980) *Table of Integrals, Series, and Products.* Orlando: Academic Press.

Hall, P. (1988) "Theoretical Comparison of Bootstrap Confidence Intervals (with discussion)." *Annals of Statistics*, 16, 927-985.

Karr, A.F. (1976) *Point Processes and Their Statistical Inference.* New York: Marcel Dekker, Inc.

Kernan, W. J., and Meeker, W. Q. (1992) "A Statistical Test to Assess Changes in Spontaneous Behavior of Rats Observed with a Computer Pattern Recognition System." *Journal of Biopharmaceutical Statistics*, 2, 115–135.

Kernan, W. J. Jr., Mullenix, P. J., Kent, R., Hopper, D. L., and Cressie, N. A. C. (1988) "Analysis of the Time Distribution and Time Sequence of Behavioral Acts." *International Journal of Neuroscience*, 43, 35–51.

Lindgren, B.W. (1976) *Statistical Theory, Third Edition.* New York: MacMillan Publishing Co., Inc.

Mullenix, P. J., Kernan, W. J., Tassinari, M. S., and Schunior, A. (1989) "Generation of Dose-Response Data Using Activity Measures." *Journal of the American College of Toxicology*, 8, 185–197.

Ohser, J. (1983) "On Estimators for the Reduced Second Moment Measure of Point Processes." *Mathematische Operationsforschung und Statistik, Series Statistics*, 14, 63–71.

Ohser, J., and Stoyan, D. (1981) "On the Second-Order and Orientation Analysis of Planar Stationary Point Processes." *Biometrical Journal*, 23, 523–533.

Ripley, B.D. (1976) "The Second-Order Analysis of Stationary Point Processes." *Journal of Applied Probability*, 13, 255-266.

Snyder, D.L. (1975) *Random Point Processes.* New York: Wiley.

Stein, M.L. (1993) "Asymptotically Optimal Estimation for the Reduced Second Moment Measure of Point Processes." *Biometrika*, 80, 443-449.

Wender, P. H. (1971) *Minimal Brain Dysfunction in Children.* New York: Wiley - Interscience.

2.9 Appendix A. Properties of the Initial Estimator

In this appendix we compute the expectation and variance of the temporal analogue to the Ohser and Stoyan (1981) estimator. First we show that $\mathrm{E}(\hat{K}(t)) = t$ and therefore is unbiased under CTR. Letting $B = \sum_{t_i,t_j} \phi(t_i,t_j)$ and using factorial moment measures as described in eqs.(2.7) and (2.8), we have

$$\mathrm{E}(\hat{K}(t)) = \frac{1}{2\lambda^2}\mathrm{E}(B),$$

where

$$\begin{aligned}
\mathrm{E}(B) &= \lambda^2 \int_{A^2} \phi(t_i,t_j)dt_i dt_j \\
&= \lambda^2 \int_0^T \int_0^T \frac{I(|t_i-t_j|<t)}{T-|t_i-t_j|}dt_i dt_j \\
&= \lambda^2 \int_0^t \left[\int_0^{t_i} \frac{1}{T+t_j-t_i}dt_j + \int_{t_i}^{t_i+t} \frac{1}{T+t_i-t_j}dt_j\right] dt_i \\
&\quad + \lambda^2 \int_t^{T-t} \left[\int_{t_i-t}^{t_i} \frac{1}{T+t_j-t_i}dt_j + \int_{t_i}^{t_i+t} \frac{1}{T+t_i-t_j}dt_j\right] dt_i \\
&\quad + \lambda^2 \int_{T-t}^{T} \left[\int_{t_i-t}^{t_i} \frac{1}{T+t_j-t_i}dt_j + \int_{t_i}^{T} \frac{1}{T+t_i-t_j}dt_j\right] dt_i \\
&= \lambda^2 \int_0^t \left[\log(T) - \log(T-t_i) + \log(T) - \log(T-t)\right] dt_i
\end{aligned}$$

$$
\begin{aligned}
&+ \lambda^2 \int_t^{T-t} [\log(T) - \log(T-t) - \log(T-t) + \log(T)]\, dt_i \\
&+ \lambda^2 \int_{T-t}^{T} \left[\log(T) - \log(T-t) - \log(t_i) + \log(T)\right] dt_i \\
= \;& \lambda^2 [2t\log(T) - t\log(T-t) + (T-t)\log(T-t) - T + t - T\log(T) + T \\
&+ 2(T-2t)\log(T) - 2(T-2t)\log(T-t) + 2t\log(T) - t\log(T-t) \\
&- T\log(T) + T + (T-t)\log(T-t) - T + t] \\
= \;& \lambda^2 2t.
\end{aligned}
$$

Thus $\mathrm{E}(\hat{K}(t)) = \frac{1}{2\lambda^2}\mathrm{E}(B) = t$.

We now proceed to calculate the variance of the estimator. The variance for B is given by eq. (2.9). We begin by calculating the pieces needed to compute (2.9).

$$
\begin{aligned}
S_1 &= \int_{A^3} \frac{I(|t_i - t_j| < t)}{T - |t_i - t_j|} \frac{I(|t_i - t_k| < t)}{T - |t_i - t_k|} dt_i dt_j dt_k \\
&= \int_0^t \left[\int_0^{t_i} \int_0^{t_i} \frac{1}{T + t_j - t_i} \frac{1}{T + t_k - t_i} dt_j dt_k \right. \\
&\quad \left. + \int_{t_i}^{t_i + t} \int_{t_i}^{t_i + t} \frac{1}{T + t_i - t_j} \frac{1}{T + t_i - t_k} dt_j dt_k \right] dt_i \\
&\quad + \int_t^{T-t} \left[\int_{t_i - t}^{t_i} \int_{t_i - t}^{t_i} \frac{1}{T + t_j - t_i} \frac{1}{T + t_k - t_i} dt_j dt_k \right. \\
&\quad \left. + \int_{t_i}^{t_i + t} \int_{t_i}^{t_i + t} \frac{1}{T + t_i - t_j} \frac{1}{T + t_i - t_k} dt_j dt_k \right] dt_i \\
&\quad + \int_{T-t}^{T} \left[\int_{t_i - t}^{t_i} \int_{t_i - t}^{t_i} \frac{1}{T + t_j - t_i} \frac{1}{T + t_k - t_i} dt_j dt_k \right. \\
&\quad \left. + \int_{t_i}^{T} \int_{t_i}^{T} \frac{1}{T + t_i - t_j} \frac{1}{T + t_i - t_k} dt_j dt_k \right] dt_i \\
&= \int_0^t \left[\left(\log(T) - \log(T - t_i)\right)^2 + \left(\log(T) - \log(T-t)\right)^2 \right] dt_i
\end{aligned}
$$

$$
\begin{aligned}
&+ \int_t^{T-t} \left[(\log(T) - \log(T-t))^2 + (\log(T) - \log(T-t))^2\right] dt_i \\
&+ \int_{T-t}^{T} \left[(\log(T) - \log(T-t))^2 + (\log(T) - \log(t_i))^2\right] dt_i \\
= \;& \int_0^t \left[(\log(T))^2 - 2\log(T)\log(T-t_i) + (\log(T-t_i))^2\right] dt_i \\
&+ t\left[\log\left(\frac{T}{T-t}\right)\right]^2 + 2(T-2t)\left[\log\left(\frac{T}{T-t}\right)\right]^2 + t\left[\log\left(\frac{T}{T-t}\right)\right]^2 \\
&+ \int_{T-t}^{T} \left[(\log(T))^2 - 2\log(T)\log(t_i) + (\log(t_i))^2\right] dt_i \\
= \;& t(\log(T))^2 + 2\log(T)[(T-t_i)\log(T-t_i) - (T-t_i)]\Big|_0^t \\
&+ \{(T-t_i)[\log(T-t_i)]^2 - 2[(T-t_i)\log(T-t_i) - (T-t_i)]\}\Big|_0^t \\
&+ 2(T-t)\left[\log\left(\frac{T}{T-t}\right)\right]^2 + t(\log(T))^2 - 2\log(T)[t_i\log(t_i) - t_i]\Big|_{T-t}^{T} \\
&+ \{t_i(\log(t_i))^2 - 2[t_i\log(t_i) - t_i]\}\Big|_{T-t}^{T} \\
= \;& 2t(\log(T))^2 + 2\log(T)[(T-t)\log(T-t) - (T-t) - T\log(T) + T] \\
&- \{(T-t)[\log(T-t)]^2 - 2(T-t)\log(T-t) + 2(T-t) - T[\log(T)]^2 \\
&+ 2T\log(T) - 2T\} + 2(T-t)\left[\log\left(\frac{T}{T-t}\right)\right]^2 - 2\log(T)[T\log(T) - T \\
&- (T-t)\log(T-t) + (T-t)] + T[\log(T)]^2 - 2T\log(T) + 2T \\
&- (T-t)[\log(T-t)]^2 + 2(T-t)\log(T-t) - 2(T-t) \\
= \;& 2t(\log(T))^2 + 2(T-t)\log(T)\log(T-t) + 2t\log(T) - 2T(\log(T))^2 \\
&- (T-t)[\log(T-t)]^2 + 2(T-t)\log(T-t) + 2t + T[\log(T)]^2 - 2T\log(T) \\
&+ 2(T-t)\left[\log\left(\frac{T}{T-t}\right)\right]^2 - 2T[\log(T)]^2 + 2(T-t)\log(T)\log(T-t) \\
&+ 2t\log(T) + T[\log(T)]^2 - 2T\log(T) - (T-t)[\log(T-t)]^2 \\
&+ 2(T-t)\log(T-t) + 2t \\
= \;& -2(T-t)[\log(T)]^2 + 4(T-t)\log(T)\log(T-t) - 4(T-t)\log(T)
\end{aligned}
$$

$$
\begin{aligned}
& -2(T-t)[\log(T-t)]^2 + 4(T-t)\log(T-t) + 2(T-t)\left[\log\left(\frac{T}{T-t}\right)\right]^2 \\
& + 4t \\
= \; & -2(T-t)\{[\log(T)]^2 - 2\log(T)\log(T-t) + [\log(T-t)]^2\} \\
& + 4(T-t)\log\left(\frac{T-t}{T}\right) + 2(T-t)\left[\log\left(\frac{T}{T-t}\right)\right]^2 + 4t \\
= \; & 4(T-t)\log\left(\frac{T-t}{T}\right) + 4t
\end{aligned}
$$

$$
\begin{aligned}
S_2 \; = \; & \int_A 2 \left(\frac{I(|t_i - t_j| < t)}{T - |t_i - t_j|}\right)^2 dt_i dt_j \\
= \; & \int_0^t \left[\int_0^{t_i} \left(\frac{1}{T + t_j - t_i}\right)^2 dt_j + \int_{t_i}^{t_i + t} \left(\frac{1}{T + t_i - t_j}\right)^2 dt_j\right] dt_i \\
& + \int_t^{T-t} \left[\int_{t_i - t}^{t_i} \left(\frac{1}{T + t_j - t_i}\right)^2 dt_j + \int_{t_i}^{t_i + t} \left(\frac{1}{T + t_i - t_j}\right)^2 dt_j\right] dt_i \\
& + \int_{T-t}^{T} \left[\int_{t_i - t}^{t_i} \left(\frac{1}{T + t_j - t_i}\right)^2 dt_j + \int_{t_i}^{T} \left(\frac{1}{T + t_i - t_j}\right)^2 dt_j\right] dt_i \\
= \; & \int_0^t \left[-\frac{1}{T} + \frac{1}{T - t_i} + \frac{1}{T - t} - \frac{1}{T}\right] dt_i \\
& + \int_t^{T-t} \left[-\frac{1}{T} + \frac{1}{T - t} + \frac{1}{T - t} - \frac{1}{T}\right] dt_i \\
& + \int_{T-t}^{T} \left[-\frac{1}{T} + \frac{1}{T - t} + \frac{1}{t_i} - \frac{1}{T}\right] dt_i \\
= \; & -\frac{2t}{T} + \frac{t}{T-t} - \log\left(\frac{T-t}{T}\right) - \frac{2(T-2t)}{T} + \frac{2(T-2t)}{T-t} - \frac{2t}{T} + \frac{t}{T-t} \\
& + \log\left(\frac{T}{T-t}\right) \\
= \; & -\frac{2T}{T} + \frac{2(T-t)}{T-t} + 2\log\left(\frac{T}{T-t}\right) \\
= \; & 2\log\left(\frac{T}{T-t}\right)
\end{aligned}
$$

These pieces are combined to obtain:

$$\begin{aligned}\text{var}(B) &= 4\lambda^3 S_1 + 2\lambda^2 S_2 \\ &= 4\lambda^3\left[4(T-t)\log\left(\frac{T-t}{T}\right)+4t\right]+2\lambda^2\left[2\log\left(\frac{T}{T-t}\right)\right]\end{aligned}$$

Then because $\text{var}(\hat{K}(t)) = \frac{1}{4\lambda^4}\text{var}(B)$,

$$\text{var}(\hat{K}(t)) = \frac{4(T-t)}{\lambda}\log\left(\frac{T-t}{T}\right)+\frac{4t}{\lambda}+\frac{1}{\lambda^2}\log\left(\frac{T}{T-t}\right).\square$$

2.10 Appendix B. Proof of Theorem 1 (Stein, 1993)

Before we prove the theorem, we need the following lemma:

Lemma 2.9.1: For $g_0(x,\phi) = 2(n-1)\text{E}[\phi(X_1,X_2)|X_1 = x]$,

$$\text{cov}_n\left(\sum_{i\neq j}\phi(X_i,X_j),\alpha(X_1)\right) = \text{cov}_n\left(\sum_{j=1}^{n} g_0(X_j),\alpha(X_1)\right), \tag{2.30}$$

for every $\alpha(\cdot)$ satisfying the following condition:

$$\int_A \alpha^2(x)dx < \infty.$$

Proof:

Taking the two sides of (2.30) separately, we see that

$$\text{cov}_n\left(\sum_{i\neq j}\phi(X_i,X_j),\alpha(X_1)\right) = 2(n-1)\text{cov}[\phi(X_1,X_2),\alpha(X_1)]$$

and

$$\begin{aligned}&\text{cov}_n\left(\sum_{j=1}^{n} g_0(X_j),\alpha(X_1)\right) \\ &= \text{cov}(g_0(X_1),\alpha(X_1)) \\ &= 2(n-1)\text{cov}[\text{E}(\phi(X_1,X_2)|X_1),\alpha(X_1)] \\ &= 2(n-1)\text{cov}[\phi(X_1,X_2),\alpha(X_1)].\end{aligned}$$

We use this result now in proving Theorem 1.

Proof:

Eq. (2.30) implies that:

$$\mathrm{cov}_n\left(\sum_{i\neq j}\phi(X_i,X_j),\alpha(X_1)\right)=\mathrm{cov}_n\left(\sum_{j=1}^{n}[g_0(X_j)-a^{-1}I_{g_0}],\alpha(X_1)\right)$$

or, that

$$\mathrm{cov}_n\left(\sum_{i\neq j}\phi(X_i,X_j)-\sum_{j=1}^{n}[g_0(X_j)-a^{-1}I_{g_0}],\alpha(X_1)\right)=0. \tag{2.31}$$

This implies that for the smaller class of functions $\{\alpha : \int_A \alpha(x)dx = 0,$ and $\int_A \alpha^2(x)dx < \infty\}$ (2.31) also holds. Let

$$\begin{aligned}\delta_0(X) &= \sum_{i\neq j}\phi(X_i,X_j)-\sum_{j=1}^{n}[g_0(X_j)-a^{-1}I_{g_0}]\\ \delta(X) &= \sum_{i\neq j}\phi(X_i,X_j)-\sum_{j=1}^{n}[g(X_j)-a^{-1}I_{g}]\end{aligned}$$

where both of these are unbiased for $\mathrm{E}(\Sigma_{i\neq j}\,\phi(X_i,X_j))$. Because

$$\delta_0-\delta=\sum_{j=1}^{n}[g(X_j)-a^{-1}I_g]-\sum_{j=1}^{n}[g_0(X_j)-a^{-1}I_{g_0}]$$

and

$$\mathrm{E}(\delta_0-\delta)=0,$$

we have eq. (2.31)

$$\mathrm{cov}(\delta,\delta_0-\delta)=0$$

and

$$\mathrm{var}(\delta_0)=\mathrm{cov}(\delta_0,\delta)\leq[\mathrm{var}(\delta_0)\mathrm{var}(\delta)]^{1/2}.$$

From this it follows that

$$\text{var}(\delta_0) \leq \text{var}(\delta).$$

We thus conclude is that δ_0 is a uniformly minimum variance unbiased estimator.

Now we compute the minimum variance, first looking at the variance and covariance of several pieces.

$$\begin{aligned}
\text{var}_n\left(\sum_{i\neq j}\phi(X_i,X_j)\right) &= n(n-1)(n-2)(n-3)a^{-4}S^2 \\
&\quad + 4n(n-1)(n-2)a^{-3}S_1 \\
&\quad + 2n(n-1)a^{-2}S_2 - n^2(n-1)^2a^{-4}S^2 \\
&= n(n-1)[(n-2)(n-3)-n(n-1)]a^{-4}S^2 \\
&\quad + 4n(n-1)(n-2)a^{-3}S_1 + 2n(n-1)a^{-2}S_2 \\
&= 2n(n-1)[(3-2n)a^{-4}S^2 + 2(n-2)a^{-3}S_1 + a^{-2}S_2].
\end{aligned}$$

$$\begin{aligned}
\text{var}[\text{E}(\phi(X_1,X_2)|X_1)] &= \text{var}\left(\int_A \frac{\phi(X_1,X_2)}{a}dX_2\right) \\
&= a^{-2}\text{var}\left(\int_A \phi(X_1,X_2)dX_2\right) \\
&= a^{-2}\int_A a^{-1}\left[\int_A \phi(X_1,X_2)dX_2\right]^2 dX_1 \\
&\quad - a^{-2}\left(\int_A\int_A \frac{\phi(X_1,X_2)}{a}dX_2dX_1\right)^2 \\
&= a^{-3}S_1 - a^{-4}S^2.
\end{aligned}$$

$$\begin{aligned}
\text{var}_n\left(\sum_{j=1}^n g_0(X_j)\right) &= \text{var}_n\left(\sum_{j=1}^n 2(n-1)\text{E}(\phi(X_j,X_2)|X_j)\right) \\
&= 4n(n-1)^2\text{var}\{\text{E}[\phi(X_1,X_2)|X_1]\} \\
&= 4n(n-1)^2(a^{-3}S_1 - a^{-4}S^2).
\end{aligned}$$

$$
\begin{aligned}
&\operatorname{cov}\left(\sum_{i\neq j}\phi(X_1,X_2),\sum_{j=1}^{n}g_0(X_j)\right)\\
&= 2(n-1)\operatorname{cov}\left(\sum_{i\neq j}\phi(X_1,X_2),\sum_{j=1}^{n}\mathrm{E}[\phi(X_j,X_2)|X_j]\right)\\
&= 2n(n-1)\operatorname{cov}\left(\sum_{i\neq j}\phi(X_1,X_2),\mathrm{E}[\phi(X_1,X_2)|X_1]\right)\\
&= 2n(n-1)\operatorname{cov}\left(\sum_{j=1}^{n}2(n-1)1\mathrm{E}[\phi(X_j,X_2)|X_j],\mathrm{E}[\phi(X_1,X_2)|X_1]\right)\\
&= 4n(n-1)^2\operatorname{cov}\{\mathrm{E}[\phi(X_1,X_2)|X_1],\mathrm{E}[\phi(X_1,X_2)|X_1]\}\\
&= 4n(n-1)^2\operatorname{var}\{\mathrm{E}[\phi(X_1,X_2)|X_1]\}\\
&= 4n(n-1)^2(a^{-3}S_1-a^{-4}S^2).
\end{aligned}
$$

Putting these pieces together we obtain the final result:

$$
\begin{aligned}
&\operatorname{var}\left(\sum_{i\neq j}\phi(X_i,X_j)-\sum_{j=1}^{n}[g_0(X_j)-a^{-1}I_{g_0}]\right)\\
&= \operatorname{var}\left(\sum_{i\neq j}\phi(X_i,X_j)-\sum_{j=1}^{n}g_0(X_j)\right)\\
&= \operatorname{var}\left(\sum_{i\neq j}\phi(X_i,X_j)\right)+\operatorname{var}\left(\sum_{j=1}^{n}g_0(X_j)\right)\\
&\quad -2\operatorname{cov}\left(\sum_{i\neq j}\phi(X_i,X_j),\sum_{j=1}^{n}g_0(X_j)\right)\\
&= 4n(n-1)(n-2)a^{-3}S_1-n(n-1)(4n-6)a^{-4}S^2+2n(n-1)a^{-2}S_2\\
&\quad +4n(n-1)^2(a^{-3}S_1-a^{-4}S^2)-8n(n-1)^2(a^{-3}S_1-a^{-4}S^2)
\end{aligned}
$$

$$
\begin{aligned}
&= 4n(n-1)[n-2+n-1-2(n-1)]a^{-3}S_1 + 2n(n-1)[3-2n-2(n-1) \\
&\quad + 4(n-1)]a^{-4}S^2 + 2n(n-1)a^{-2}S_2 \\
&= 2n(n-1)[a^{-4}S^2 - 2a^{-3}S_1 + a^{-2}S_2].\square
\end{aligned}
$$

2.11 Appendix C. Properties of Temporal Analogue to Ohser and Stoyan's (1981) $\phi(x,y)$.

We now demonstrate that the four conditions stated in Section 2.3.2 hold for this function. The first condition states that $\phi(x,y)$ is uniformly bounded as T increases. For any given realization of this function, the constant t must be given and in fact it must be less than T. Start with a minimum T_0, (i.e. a minimum observation region $[0,T_0]$), and let t be some constant satisfying $t\epsilon[0,cT_0]$ where $0 < c < 1$. Then for every x, y, and $T \geq T_0$,

$$
\begin{aligned}
\phi(x,y) &= \frac{TI(|x-y| < t)}{2(T-|x-y|)} \\
&= \frac{T}{2(T-t)} \\
&\leq \frac{T_0}{2(T_0 - cT_0)} = \frac{1}{2(1-c)}.
\end{aligned}
$$

Therefore $\phi(x,y)$ is uniformly bounded as T grows.

The second condition states that the function should be 0 for $|x-y| > t$. The indicator function in the numerator of $\phi(x,y)$ assures that this condition holds.

The third condition requires that $\phi(x,y)$ be measurable with respect to $|x-y|$. But indicator functions are measurable, so the result follows.

Finally, $\phi(x,y)$ must satisfy (2.19). That is

$$\frac{1}{T}\int_A [\phi(x,x+r)I(x+r\epsilon A)+\phi(x,x-r)I(x-r\epsilon A)]dx = I(r<t). \qquad (2.32)$$

Substituting

$$\phi(x,y) = \frac{TI(|x-y|<t)}{2(T-|x-y|)}$$

into the left hand side of (2.32) gives

$$\begin{aligned}
&\frac{1}{T}\int_0^{T-r}\frac{TI(r<t)}{2(T-r)}dx + \frac{1}{T}\int_r^T \frac{TI(r<t)}{2(T-r)}dx \\
&= \frac{1}{T}\frac{TI(r<t)}{2(T-r)}(T-r) + \frac{1}{T}\frac{TI(r<t)}{2(T-r)}(T-r) \\
&= I(r<t).
\end{aligned}$$

Therefore, the given $\phi(x,y)$ satisfies all four conditions stated in Section 2.3.2.

2.12 Appendix D. Properties of the Final Estimator

In this appendix we compute the variance of the final estimator, discussed in Section 2.3.3.

Since this estimator was obtained by using Theorem 1, this theorem can also be used to calculate its variance. Eq. (2.14) gives the form of the variance as:

$$\text{var}\left(\frac{n^2}{T}\hat{K}(t)\right) = 2n(n-1)(T^{-2}S_2 - 2T^{-3}S_1 + T^{-4}S^2),$$

conditional on n. This means that

$$\text{var}_n(\hat{K}(t)) = \frac{T^2}{n^4}2n(n-1)(T^{-2}S_2 - 2T^{-3}S_1 + T^{-4}S^2).$$

We proceed by calculating S, S_1, and S_2.

$$S = \frac{T}{2}\int_{A^2}\frac{I(|t_j-t_i|<t)}{T-|t_j-t_i|}dt_i dt_j$$

$$
\begin{aligned}
&= \frac{T}{2}\int_0^t \left[\int_0^{t_i} \frac{1}{T+t_j-t_i}dt_j + \int_{t_i}^{t_i+t} \frac{1}{T+t_i-t_j}dt_j\right] dt_i \\
&\quad + \frac{T}{2}\int_t^T -t\left[\int_{t_i-t}^{t_i} \frac{1}{T+t_j-t_i}dt_j + \int_{t_i}^{t_i+t} \frac{1}{T+t_i-t_j}dt_j\right] dt_i \\
&\quad + \frac{T}{2}\int_T -t^T\left[\int_{t_i-t}^{t_i} \frac{1}{T+t_j-t_i}dt_j + \int_{t_i}^{T} \frac{1}{T+t_i-t_j}dt_j\right] dt_i \\
&= \frac{T}{2}\int_0^t [\log(T)-\log(T-t_i)+\log(T)-\log(T-t)]dt_i \\
&\quad + \frac{T}{2}\int_t^{T-t} [\log(T)-\log(T-t)-\log(T-t)+\log(T)]dt_i \\
&\quad + \frac{T}{2}\int_{T-t}^{T} [\log(T)-\log(T-t)-\log(t_i)+\log(T)]dt_i \\
&= \frac{T}{2}\{2t\log(T)-t\log(T-t)+(T-t)\log(T-t)-T+t-T\log(T)+T \\
&\quad + 2(T-2t)\log(T)-2(T-2t)\log(T-t)+2t\log(T)-t\log(T-t) \\
&\quad - T\log(T)+T+(T-t)\log(T-t)-T+t\} \\
&= tT.
\end{aligned}
$$

Next, to compute S_1 and S_2 notice that the form of $\phi(\cdot,\cdot)$ used in this estimator is simply the version of $\phi(\cdot,\cdot)$ used in the initial estimator multiplied by $\frac{T}{2}$. This means that we can take both the S_1 and S_2 computed in Appendix A and multiply them by $\frac{T^2}{4}$. This gives:

$$
\begin{aligned}
S_1 &= T^2(T-t)\log\left(\frac{T-t}{T}\right) - tT^2, \\
S_2 &= \frac{T^2}{2}\log\left(\frac{T}{T-t}\right).
\end{aligned}
$$

We can now compute the variance of $\hat{K}(t)$, conditional on n, as:

$$
\mathrm{var}_n(\hat{K}(t)) = \frac{2T^2(n-1)}{n^3}\left\{\frac{1}{T^2}\frac{T^2}{2}\log\left(\frac{T}{T-t}\right)\right.
$$

$$
\begin{aligned}
& - \frac{2}{T^3}\left[T^2(T-t)\log\left(\frac{T-t}{T}\right) - tT^2\right] \\
& + \frac{1}{T^4}t^2T^2\Big\} \\
= \; & \frac{2T^2(n-1)}{n^3}\left[\left(\frac{1}{2} - \frac{2(T-t)}{T}\right)\log\left(\frac{T}{T-t}\right) + \frac{2tT+t^2}{T^2}\right].\square
\end{aligned}
$$

3. STATISTICAL TESTS FOR SIGNALS IN CATEGORICAL TEMPORAL DATA

A paper to be submitted to the *Biometrical Journal*

Ann Cannon and William Meeker[1]

3.1 Abstract

In an attempt to discover differences in behavior between two groups of rats, one treated and one control, many different correlated test statistics are computed. Previously, a decision was made by determining which of the test statistics were significant. Kernan and Meeker (1992) created a single test statistic by counting the number of significant individual tests. This article goes further, giving two possible alternatives to the Kernan and Meeker (1992) statistic: the sum of the squared test statistics and a Wald-like combination of the test statistics using the covariance matrix. The asymptotic null distributions of all three statistics are given, as well as a method for computing simulated distributions using the bootstrap method. The use of all three methods are then demonstrated on each of three data sets. Finally, a simulated power study reveals that the Wald-like statistic is much better than the other two, leading to the suggestion of its use in place of the other two statistics.

[1]Principal Author is Ann Cannon. William Meeker provided guidance throughout the project.

3.2 Introduction

Over the years, several very effective medicinal drugs have been found to have serious side effects, often discovered long after the primary cure has been effected. Some of these side effects might have been detected by behavioral differences in the patient although the side effect is most likely physiological. Researchers have designed experiments using rats as an animal model to try to predict the possibility of a side effect measurable by a behavior change for drugs developed for humans. In this article we discuss such an experimental design and the analysis of the data received from it. We compare the current analysis technique with two new methods that we propose.

3.2.1 Motivating experiment

The specific design of the experiment is as follows (e.g., Kernan, Mullenix, Kent, Hopper and Cressie, 1988): pairs of rats were chosen, one in each pair injected with a saline solution, the other with the experimental drug. At some later point in time these pairs were observed as they explored a novel environment and their act or position at the beginning of each of the 900 consecutive seconds was recorded. In earlier experiments, the experimenters used 900 still-frame photographs and visual discrimination to code the photographs; more recently, experimenters have used television cameras and a computer pattern recognition system to do the coding. Observed behavioral differences between the rats provide evidence of physiological change.

The data from this experiment was coded as follows. The behavior of the rat at each of the 900 time points was categorized into one of five body positions: standing, sitting, rearing, walking, and lying down; and one of the eight modifiers: blank (no

recognized activity), groom, head turn, turn, look, smell, sniff, and washing face. The combined set of 5 body positions and 8 modifiers are defined to be the set of 13 regular acts. Norton, Mullenix and Culver (1976) reclassified each of the regular acts into one of three classes labeled "grooming," "exploratory," and "attention." Since each of the 900 seconds consists of two regular acts, both a body position and a modifier, Kernan and Meeker (1992) mapped the combination of body position and modifier at each second into one of the six combinations of two classes, e.g., "attention-attention," "attention-explore," "groom-attention." It should be noted that these new combinations did not reflect which classification each specific regular act had separately, merely the two classes that the two regular acts together represented. These six combinations are labeled the combined acts. Kernan and Meeker (1992) then used the information from both the regular acts (of which there are two for each time period) and the combined acts (of which there is one for each time period) in their analysis of the data.

The experiment described above leads to a raw data set with observations on $2n$ rats, where n denotes the number of pairs observed. Each observation consists of the following information: a pair number, a treatment/control indicator variable, and, for each of the 900 discrete time periods, the values of the two categorical response variables that represent the body position and modifier, respectively, of the act that the rat performed. From these data the goal is to detect systematic behavioral differences that exist between the treated and the control rats. At this stage the amount of difference is not of primary concern, merely whether a difference, not due to usual rat-to-rat variability, exists or not. If a systematic difference is found, the researchers would use this as evidence that some physiological change may have taken

place and more testing would be done to find the particular physiological difference in question. At the outset of these experiments, the researchers often do not have an idea about what type of systematic behavioral difference might exist, so they look for any detectable change.

3.2.2 History of the analysis from this type of experiment

Several ways of quantifying data from this type of experiment have evolved. Originally, researchers simply counted the number of times that each act was initiated (Kernan et al., 1988). Psychiatrists, however, found that the total time a rat performed an act, the average duration of the acts, and the time lapsing between initiations of an act were also of interest (Pohl, 1976; Baumeister, 1978). Wender (1971) also showed that some types of behaviors, such as hyperactivity in children, will not be discovered using only the number of initiations, the total time, the average duration and the time lapsing between initiations. Hyperactivity has been traditionally seen as a state of increased motor activity, but this sense of increase comes from the behavior occurring in a different pattern than normal rather than a true increase in activity level. An unexpected behavior pattern may lead to the appearance of more activity while the average amount of time spent doing each act is really the same as in a normal child (Pontius, 1973). In order to detect a difference in the pattern of the behavior from the expected pattern of behavior, Kernan et al. (1988) looked more into the structure of the observations, using a temporal analogue of the spatial statistic, the K-function, to define what they called the time distribution of the acts and the time sequence of pairs of acts.

Kernan, Mullenix and Hopper (1989) went one step further by looking at many

test statistics aimed at describing differences between the groups for various aspects of the data, hoping that by looking at many statistics, any possible behavioral difference might be detected. The researchers then based their conclusions on which statistics were found to be significant (Kernan, Mullenix and Hopper, 1989). The Kernan, Mullenix and Hopper (1989) method evolved into a new overall test statistic, consisting of the ratio of the number of individual test statistics that were found to be significant to the number of tests done (Kernan and Meeker, 1992). By using a jackknife-type simulation method with control-control data, Kernan and Meeker (1992) computed the null distribution for this overall statistic for experiments consisting of 17 and 20 pairs of rats.

3.2.3 Analysis of similar experiments

Others have also approached the problem of analyzing many possibly correlated test statistics. For instance, Westfall and Young (1989) approach the question of multiple tests by using resampling techniques to adjust the p-values (generally upward), considering both the number of tests being performed and the correlational structure between the test statistics, before determining which tests are significant. Their techniques are appropriate for the case of determining which of the individual tests is significant while still dealing with the multiplicity of the tests. They report that their procedure will have less power than global tests when there is a mild departure from the null for many tests, but that it will do better than global tests when there are great differences for a few of the individual tests. In our problem, the researchers are not yet interested in *what* the differences are, merely whether any exist. So for this stage, we believe that a global statistic like the ones we are about to suggest

make more sense. However, after it has been determined that a difference exists, the p-value adjustment may make sense in the second stage of determining the particular nature of the difference.

Haccou and Meelis (1992) concern themselves with the same specific type of data set considered in this article, vectors of animal behavior data. They use a third approach to analyzing this type of data, namely a continuous time Markov-Chain model. Because they found so much animal-to-animal variability, Haccou and Meelis consider each animal as a separate data set and as being a realization of a (possibly) different stochastic process (pg. 4). Therefore their techniques are applicable, as outlined in Haccou and Meelis (1992), only for comparing two realizations, rather than two groups of realizations. We believe that under all of the variability lies the same signal (model) for each of the rats within a group and therefore have combined the information into one comparison.

In Section 3.3 we outline in detail methods suggested by Kernan and Meeker (1992). We go on, in Section 3.4, to suggest two new statistics, giving asymptotic results for both the two new statistics and the one already in use, suggested by Kernan and Meeker (1992). Section 3.5 contains numerical results of the three statistics for three different data sets. Section 3.6 contains simulated distributional results of the three statistics and Section 3.7 is a comparison of the simulated power of all three statistics, from which we recommend the use of one particular statistic. In Section 3.8 we discuss our results and suggest areas for future research.

3.3 Recent Developments for Data from this Type of Experiment

Kernan and Meeker (1992) started by computing four types of individual statistics for both the set of combined acts as well as the set of regular acts. These four statistics are:

1. *Number of Initiations.* An initiation of an act was defined to be the first frame of the one or more consecutive frames where the animal performed the specific act.

2. *Total Time.* The total number of frames where the act occurred.

3. *Time Distribution.* A statistic describing the pattern of where an act occurs over time.

4. *Time Sequence of Pairs of Acts.* A statistic describing the pattern of how a pair of acts occurs over time.

These last two types of statistics are based on estimating the temporal analogue to the spatial statistic, the K-function. Kernan and Meeker (1992) used an estimator of the K-function derived in Kernan, et al. (1988) to define both the time distribution and the time sequence of pairs of acts for the two groups of rats. The estimator for the temporal K-function in Kernan et al. (1988) has been improved upon and used in the development of a new estimator for the time distribution in Cannon and Cressie (1993) and this new estimator will be used for the analysis presented here. The related statistic for the time sequence of pairs of acts (based on the cross K-function), using the new estimator for the K-function, has not been developed yet and so we will use only the first three types of statistics in the analysis presented in

Sections 3.4 through 3.7. Once a K-function statistic for the time sequence of pairs of acts has been developed, it will be possible to include them in an analysis by using the results from this article.

3.3.1 K-functions

Both the number of initiations of an act and the total time spent performing an act are straightforward to calculate. The time distribution, however, is complicated enough that we explain its background, meaning and use. The K-function, sometimes called the reduced second moment measure, gives a quantification of the spatial dependence between different regions of a stationary point process. For the type of experiment analyzed in this work, it illustrates the temporal dependence of the initiations of an act. The K-function in two- or three-dimensional space is defined as:

$$K(h) \equiv \lambda^{-1} E(\text{number of extra events within distance h of an arbitrary event})$$

(e.g., Diggle, 1983, p.47). In time this becomes:

$$K(t) \equiv \lambda^{-1} E(\text{number of extra events occurring within time } t \text{ } \textit{after} \text{ an arbitrary event.})$$

These functions are used to identify a structure in the pattern of initiations of an act over time. It is possible to find the same number of initiations and the same amount of time spent performing a specific act for treated animals as for control animals, but find that the pattern of the initiations of this act to be different. For example, one type of animal may have seemingly random initiations of an act where the other might display a nonrandom pattern. One use of K-functions is in the

calculation of the time distribution of an act. This statistic is used to find any time dependence or differences in time dependence that might exist for the initiations of an act.

The time distribution, as given by Cannon and Cressie (1993), is:

$$\begin{aligned}\hat{K}(t) \;=\; \frac{T}{n^2}\Bigg\{ & \sum_{t_i \neq t_j} \frac{TI(|t_i - t_j| \leq t)}{2(T - |t_i - t_j|)} \\ & - (n-1)\Bigg[2n\log(T) - n\log(T-t) - \sum_{0<t_i<t} \log(T - t_i) \\ & - \sum_{t \leq t_i \leq T-t} \log(T-t) - \sum_{T-t<t_i\leq T} \log(t_i) - n\frac{2t}{T}\Bigg]\Bigg\}. \qquad (3.1)\end{aligned}$$

In this equation n denotes the number of initiations of the act and T denotes the total time over which the analysis is performed. This time, T, is, however, not the same as the total time that the rats were observed. The K-function analysis is performed only on the initiations of the act. The duration of time that the rats performed the act continuously is not of interest here. Thus we reduce the original data vector somewhat by eliminating the continuations of the act of interest after each of its initiations. T represents the length of this reduced data vector. Also, t_i and t_j are the ith and jth occurrences of the act, respectively, and $I(|t_i - t_j| < t)$ is 1 (or 0) according to whether the pair (t_i, t_j) of initiations of the act occurred (did not occur) within a time separation t. $\hat{K}(t)$ is evaluated for every integer value of t between 1 and 100 to get a full picture of the time pattern of the particular act in question. Evaluating (3.1) for small values of t gives the pattern of initiations on a small scale, detecting whether points close together are clustered, regular or random. Large values of t are needed to see the macro pattern of the initiations. Cannon

and Cressie (1993) developed paired test statistics based on the areas between these functions and the theoretical K-function under the hypothesis of Complete Temporal Randomness. These statistics were found to be approximately normally distributed, at least in part, because of the central limit theorem. Cannon and Cressie (1993) also developed similar test statistics to compare a set of pairs of these functions; one in each pair coming from a treated rat, one from a control rat.

Kernan and Meeker (1992) placed a restriction on the use of of the K-functions in order to limit consideration to those statistics for which there was sufficient information. Their rule for the calculation of the K-functions was to use only those acts that had an average number of initiations per animal of ten or greater for both the control group and the exposed group separately. They believed that acts with fewer initiations would not bring enough accurate information into the function. We continue to use this restriction.

3.3.2 Kernan and Meeker (1992) approach for obtaining a single test statistic

First each of the four statistics defined at the beginning of Section 3.3 had to be computed for each of the regular and combined acts (or pairs of acts) defined in Section 3.2. Because of the restriction on the use of K functions described above, the body position "lying down" was never used in the computation of the statistics "time distribution" or "time sequence". Therefore there were 13 "total time" statistics, 13 "number of initiations" statistics, a maximum of 12 "time distribution" statistics and a maximum of 68 "time sequence" statistics computed using the 13 regular acts (5 body positions and 8 modifiers) for a maximum of 106 statistics. There were also

6 "total time" statistics, 6 "number of initations" statistics, a maximum of 6 "time distribution" statistics, and a maximum of 30 "time sequence" statistics using the 6 combined acts for a maximum of 48 statistics. The resulting maximum number of statistics for a data set is 154. Some of these statistics may not be calculated for every experiment of this type because of the restrictions mentioned above, but a large majority will be used. At this point, then, there were approximately 154 statistics to use in answering the question "Is there a difference between the two groups?" Kernan and Meeker (1992) suggested two ways of combining the statistics to help answer the question. First they created a vector $\mathbf{X}$, of length at most 154, containing the statistics that can be used to test for differences between the two groups. Then $\mathbf{X}$ was mapped into a vector $\mathbf{Y}$ of 0's and 1's where a 1 signified that the corresponding element of $\mathbf{X}$ showed a statistically significant difference.

At this point the analysis took into consideration the fact that each observation of the animals was used twice, once in the computation of the statistics for the regular acts and again for the combined acts. Because of this overlap, Kernan and Meeker (1992) divided their vector $\mathbf{Y}$ into two parts, $\mathbf{Y}_1$ based on the regular acts and $\mathbf{Y}_2$ based on the combined acts. They designated S_1 and S_2 as the sums and T_1 and T_2 as the lengths of these two vectors respectively. In their example $T_1 \leq 106$ and $T_2 \leq 48$. The two statistics that Kernan and Meeker (1992) suggested are

$$RTOT = \frac{S_1 + S_2}{T_1 + T_2}$$

and

$$RSQR = \left(\left(\frac{S_1}{T_1} \right)^2 + \left(\frac{S_2}{T_2} \right)^2 \right)^{\frac{1}{2}}.$$

Because of the structure of the actual data analysis, T_1 will always be larger than

T_2 which results in $RTOT$ more heavily weighting the information from the regular acts. $RSQR$ weighs each of the subvectors more evenly, possibly a desirable condition since each subvector, in some sense, summarizes the data set.

Kernan and Meeker (1992), using previous control-control and control-treatment data and jackknife techniques, explored the distributional characteristics of both statistics. They found that there was an extremely high degree of correlation between $RSQR$ and $RTOT$, indicating that either could be used. By using Monte Carlo calculations on a data set from a control-treatment experiment where a systematic behavior difference was known to exist, they showed that these statistics can be useful in determining whether a systematic behavioral difference exists or not. The simulation results presented in Kernan and Meeker (1992), however, only give procedures for experiments with 17 or 20 pairs of animals. They suggest that one might be able to interpolate, using their results, for 18 or 19 pairs, under a similar experimental design, but their work does not directly apply to other numbers of pairs. For experiments of other sizes more simulations would have to be run.

3.4 Alternative Statistical Methods

3.4.1 Background

In essence Kernan and Meeker (1992) performed a number of two-sample hypothesis tests and counted the number of tests that were found to be statistically significant. This count was used to determine the answer to the question "Is there a statistically significant difference between the two groups of rats?" Because the sampling distribution of the overall statistic was not known, Kernan and Meeker (1992) used a jackknife-like resampling procedure to obtain critical values for the overall

test statistic. In this section we propose two alternatives to the Kernan and Meeker (1992) overall statistic. Here and in subsequent sections we will compare the three methods.

Before proposing the new statistics, however, we suggest one change in procedure from Kernan and Meeker (1992). They used two-sample t tests, because the pairs of rats were chosen primarily for convenience of observation, as discussed in Section 3.2.1. There is, however, a potential for correlation within pairs because the rats within pairs were observed together. This experimental design allows for possible uncontrollable differences between pairs. Although in the data sets used by Kernan and Meeker (1992) the within pair correlation did not seem to exist, the potential for such correlation does exist so we suggest using paired tests instead of two-sample tests.

3.4.2 Test statistics defined

Let $t_{i,n}$, $i = 1, ..., m$ be m statistics, each based on samples of size n and let $\underline{t} = (t_{1,n}, ..., t_{m,n})'$.

3.4.2.1 The Kernan and Meeker (1992) statistic. The Kernan and Meeker (1992) statistic can be defined as:

$$S_B = \sum_{i=1}^{m} I\left(|t_{i,n}| > t_{1-\alpha/2,n-1}\right) \tag{3.2}$$

where I is the indicator function and $t_{1-\alpha/2,n-1}$ is the $1-\alpha/2$ quantile of the t distribution with $n-1$ degrees of freedom.

3.4.2.2 A statistic based on the sum of squares of $\underline{t}$. The first comparison statistic we suggest is the sum of the squared test statistics computed from the sample (which can also be thought of as a weighted sum of the original observed summary statistics with the weights being the individual standard deviations.). Define the first comparison statistic as:

$$S_T = \sum_{i=1}^{m} t_{i,n}^2. \tag{3.3}$$

This statistic, while still rather simple to calculate, will likely carry with it more information than S_B because S_T does not dichotomize the original information, and will, therefore, take into account the size of any departures from the individual hypotheses. However, S_T, like S_B, does not take into account the correlational structure between the test statistics. Because of this limitation, we suggest another possible statistic which is somewhat more complicated but which takes the correlational structure of the test statistics into account.

3.4.2.3 A statistic based on a quadratic form in $\underline{t}$. The second statistic we suggest is a Wald-like combination of the individual test statistics. To define this statistic, let $\mathbf{\Sigma}$ be the true covariance matrix associated with this set of test statistics and let $\hat{\mathbf{\Sigma}}$ be the estimated sample covariance matrix and note that its spectral decomposition into its eigenvectors and eigenvalues is

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{Q}}\hat{\mathbf{\Lambda}}\hat{\mathbf{Q}}'$$

where $\hat{\mathbf{Q}}$ is the matrix of normalized eigenvectors and $\hat{\mathbf{\Lambda}}$ is a diagonal matrix of the eigenvalues. Because the sample covariance matrix must be nonnegative definite, all eigenvalues are greater than or equal to 0 and the rank of the matrix is the number

of positive eigenvalues. For the problem at hand, $\mathbf{\Sigma}$ will usually not be of full rank and we will assume that the rank is known to be r. Partition the matrices $\hat{\mathbf{Q}}$ and $\hat{\mathbf{\Lambda}}$ as

$$\hat{\mathbf{Q}} = \begin{pmatrix} \hat{\mathbf{Q}}_1 & \hat{\mathbf{Q}}_2 \end{pmatrix},$$
$$\hat{\mathbf{\Lambda}} = \begin{pmatrix} \hat{\mathbf{\Lambda}}_1 & 0 \\ 0 & 0 \end{pmatrix},$$

where $\hat{\mathbf{Q}}_1$ is of dimension $p \times r$ and $\hat{\Lambda}_1$ is of dimension $r \times r$. Then, define

$$\hat{\mathbf{A}} = \hat{\mathbf{Q}}_1 \hat{\mathbf{\Lambda}}_1^{-1} \hat{\mathbf{Q}}_1'.$$

Finally, let

$$S_W = \underline{t}' \hat{\mathbf{A}} \underline{t}. \tag{3.4}$$

The S_W statistic, while somewhat more complicated, takes the correlations between the original t-statistics into consideration. One result of using the correlations in the computation of the test statistics is that, for example, the statistic would treat the case where two non-correlated t's are both significant, differently than if two highly correlated t's are both significant. Neither S_B nor S_T would distinguish between these two cases. The drawback to S_W is that much more computation is involved and typically the matrix $\mathbf{\Sigma}$ is unknown and must be estimated from the available data.

In Section 3.4.3 we discuss the computation of the statistics including the estimation of $\mathbf{\Sigma}$ where necessary, the simulation of the distribution of the test statistics, and the computation of the critical values for hypothesis tests using the test statistics. In Section 3.5 we compare the three statistics to each other for three different data sets.

3.4.3 Computation of the test statistics

For a given data set the actual computations of the test statistics S_B and S_T are straightforward following equations (3.2) and (3.3). S_W requires more work because an estimate of the covariance matrix for the test statistics is required. We begin by discussing how to find an estimate of Σ which will be used both in the computation of S_W and in finding the critical values for all three test statistics.

3.4.3.1 Computing the estimated covariance matrix. In order to estimate the covariance between any two test statistics we need to have a substantial number of realizations of those two statistics. This could be done by conducting a sequence of complete experiments run under the same conditions, an event that is not likely to occur in practical applications. We can, however, approximate this process by employing the bootstrap method on the data set. For this purpose the bootstrap is implemented by choosing, with replacement, samples of rat pairs of size 20 (the original sample size) from the original sample of rat pairs and recomputing the test statistics for each new bootstrap sample. In this way we create a collection of "re-samples" of the test statistics from which we can estimate the covariance matrix of the test statistics. The empirical covariance matrix computed in this manner will converge in conditional probability to the theoretical covariance matrix of the test statistics. For proof of that last statement, see Appendix A.

For our purposes, we used 300 bootstrap samples to compute the covariance matrix. Although a larger number of bootstrap samples might be a little better, larger bootstrap sample sizes would add proportionately to the amount of computer time required for our simulations. We have chosen relatively small bootstrap sample

sizes for estimating the covariances because we have seen that the amount of Monte-Carlo variance present does not affect the estimates very much. Later, when we use the bootstrap to estimate confidence levels and percentiles of simulated distributions we use a much larger number of bootstrap samples, because, for these purposes the Monte-Carlo variation does make a large difference.

3.4.3.2 Computation of S_W. The statistic S_W depends on the covariance matrix only through the positive eigenvalues and the corresponding normalized eigenvectors. So, once the estimated covariance matrix has been found, the eigenvalues and eigenvectors of $\hat{\Sigma}$ must be computed and the number of positive eigenvalues determined. Typically, with matrices as large as we are dealing with, a computer is used to compute the eigenvalues and eigenvectors and rarely will any of the eigenvalues be found to be exactly equal to 0. It is clear that any eigenvalue that is of the order of 10^{-10} or less can be safely assumed to be 0. But what about an eigenvalue on the order of 10^{-2} or 10^{-4}? We suggest looking at a scree graph of the eigenvalues (e.g., Cattell (1966)) which is a plot of the eigenvalues in descending order. Figure 3.1 is such a plot for the EC-1 data set considered in Section 3.5. To determine the number of non-zero eigenvalues, we look for the last visible drop from one eigenvalue to the next and use all eigenvalues up to the point of that last drop. Of course this decision will depend on the resolution of the graph, but the results from the methods we present are not very sensitive to the exact number of positive eigenvalues chosen. After this point, the remaining eigenvalues should form an approximately horizontal line located at zero. For this particular data set we chose to use the first 20 of the eigenvalues and eigenvectors in the computation of S_W. For the other two data sets

discussed in Section 3.5, we chose 19 eigenvalues in the same manner.

3.4.3.3 Computing the critical values for the three test statistics. To find the critical values for testing each of the three statistics, we rely on simulation of the distributions for the test statistics because the actual (non-asymptotic) distributions are not readily available. We show in Section 3.6 that the asymptotic distributions, while close, are not completely accurate representations of the test statistics computed from samples with the finite sample size of 20 as in the experiment described above.

Simulated distribution of S_B and S_T. The simulation process starts with simulating at least 2000 sets of the correct number of independent statistics (in our data sets, 26 statistics with a t_{19} distribution and 18-20 statistics with a standard normal distribution, depending on the data set used). We chose to use a simulated distribution of length 2000 because we will be using this distribution to compute critical values for hypothesis tests. With simulations of 2000, the upper 5% point was quite stable. But for critical values for tests with smaller values of α, a large simulation should be run. Let $\underline{t}_i^*$ denote one of the 2000 vectors of simulated independent statistics and T^* be the matrix with the $\underline{t}_i^*$ in the columns. Then, to create samples of statistics which have the same approximate covariance structure as the data set in question, we standardize the statistics to have variance 1 (this means dividing the t statistics by 1.06, and doing nothing to the normal statistics). At this point we have 2000 sets of statistics which have the identity matrix, $\mathbf{I}$, as their theoretical covariance matrix. In order to change these statistics so that they have a covariance matrix that is approximately $\hat{\mathbf{\Sigma}}$, we pre-multiply each set of standardized

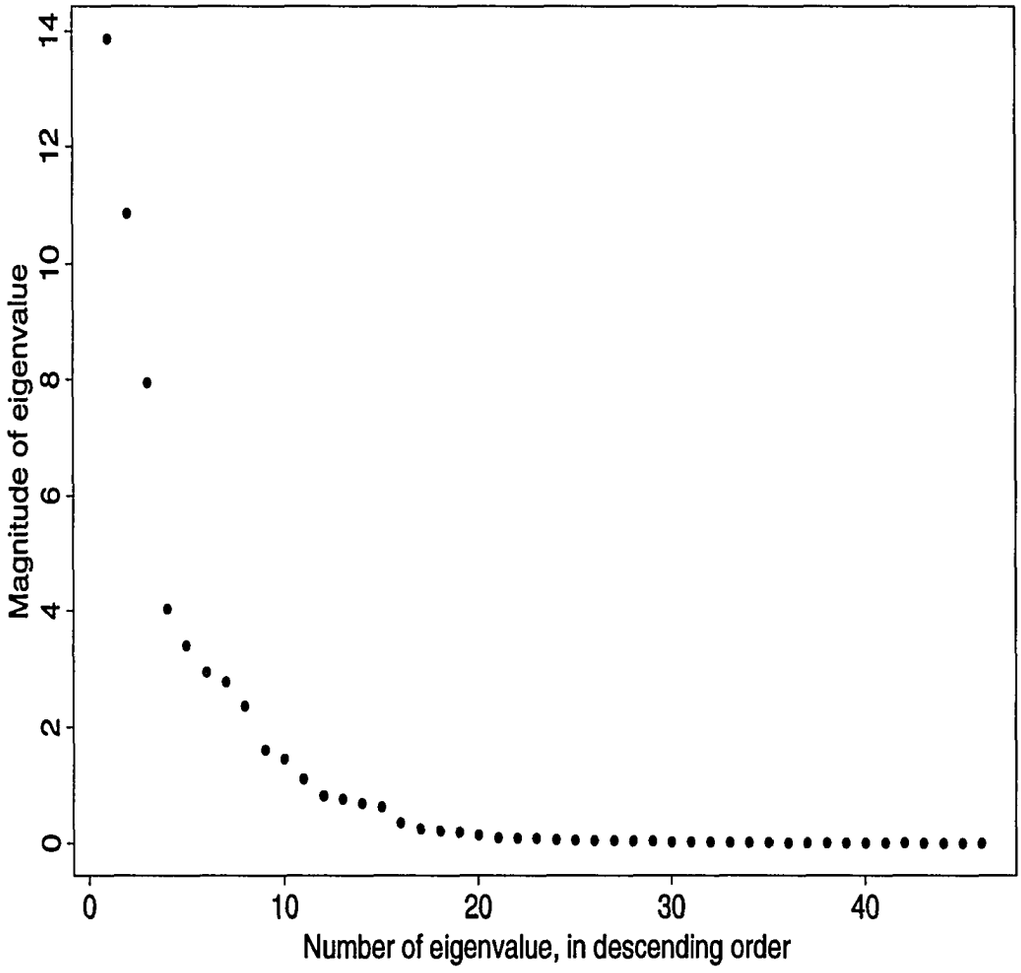

Figure 3.1: Eigenvalues of EC-1 data set plotted in descending order.

statistics by the transpose of the Choleski decomposition of the estimated covariance matrix of the data set. The Choleski decomposition gives a matrix $\mathbf{S}$ such that $\mathbf{S}'\mathbf{S} = \hat{\mathbf{\Sigma}}$. So, given that

$$\text{var}(\underline{t}^*) = I,$$

$$\text{var}(\mathbf{S}'\underline{t}^*) = \mathbf{S}'I\mathbf{S} = \hat{\mathbf{\Sigma}}.$$

The result is a set of 2000 simulated samples of statistics which have been sampled from a set of statistics having the same covariance structure as the estimated covariance structure of the original data. We use this simulated set of statistics to compute 2000 simulated values of both S_B and S_T. We now have approximate distributions for S_B and S_T that can be used to find the critical value for the hypothesis test of interest. If α is chosen to be less than .05, more simulations should be run.

Simulated distribution of S_W. The final simulation to consider is that of S_W. Simulating a set of 2000 values for S_W is not as straightforward as the process described above for S_B and S_T because S_W depends on an estimate of $\mathbf{\Sigma}$. In order to account for the variability introduced by estimating $\mathbf{\Sigma}$ in the distribution of S_W, a different estimate of $\mathbf{\Sigma}$ is needed for each simulation sample. Recall, now, that the original estimate of $\mathbf{\Sigma}$ was computed using a bootstrap method on the data set. It is not possible to use the bootstrap method on the simulation sample because we have no "raw data" from which to resample and recompute the test statistics. Instead, for each value in the simulated distribution we will simulate a set of statistics which have a covariance matrix which is approximately the same as the data set of interest. Call the vector of simulated statistics $\underline{t}$. Then we will re-estimate the covariance matrix of the data set and use the eigenvalues and eigenvectors of the new estimated

covariance matrix along with $\underline{t}$ from above to compute S_W. The complete algorithm for computing the simulated distribution of S_W is as follows:

1. Compute an estimate of the covariance matrix of the test statistics:
   (a) Select a bootstrap sample of 20 pairs of rats chosen with replacement.
   (b) Compute the statistics which do not use the K function.
   (c) Compute a K-function statistic:
      i. Compute the non-standardized K-function statistic
      ii. Select a bootstrap sample of 20 pairs of rats chosen with replacement from the 20 pairs in the sample.
      iii. Compute the non-standardized K-function statistic of the bootstrap sample.
      iv. Repeat steps (ii) and (iii) B_1 times.
      v. Use the non-standardized K-function statistics from the B_1 bootstrap samples to compute the variance of the K-function statistics.
      vi. Use the variance computed in step (v) and the statistic computed in step (i) to compute the standardized K-function statistic.
   (d) Repeat step (c) for each K-function statistic to be used.
   (e) Save the vector of test statistics for the bootstrap sample.
   (f) Repeat steps (a) through (e) B_2 times to get B_2 vectors of test statistics.
   (g) Use the B_2 vectors of test statistics to compute an estimated covariance matrix.

2. Simulate a vector of test statistics under the null hypothesis using the original covariance matrix computed the first time through step (1) as the covariances among the test statistics.

3. Repeat step (1) to compute a new estimate of the covariance matrix.

4. Combine the products of step (2) and step (3) to compute a value of S_W for the simulated distribution.

5. Repeat steps (2)-(4) B_3 times to estimate the simulated distribution of S_W.

In our computation of the simulated distribution we used $B_1 = B_2 = 300$ and $B_3 = 2000$. Notice that step (1) is the original bootstrap technique described in Section 3.4.3.1. For each of the $B_3 = 2000$ covariance estimates a bootstrap sample of size $B_2 = 300$ is required. But for this particular type of data set, where K-functions are being used, a third level of the bootstrap of size $B_1 = 300$ is required. Each statistic based on the K-function estimate requires a variance estimate computed using yet another application of the bootstrap (Cannon and Cressie, 1993). We have limited the inside two levels of the bootstrap to 300 each. The computer time required for more samples becomes prohibitively large.

We can now find the critical value needed to test the appropriate hypothesis from this simulated distribution of S_W. Again, if $\alpha < .05$ is chosen, more simulations should be run.

3.4.3.4 Finding the critical values.

Finding the critical value for S_B. Care has to be taken when computing the critical value of S_B, as the statistic has a discrete distribution; we cannot

simply choose α to be, say .05, and find the value of S_B at the beginning of the upper 5 percent point of the distribution. For example, the 5 percent point would be the 1900th point of the sorted distribution assuming a simulated distribution of length 2000. Because S_B is discrete the 1900th point of the simulated distribution gives size equal to .05 only if the 1899th position is a smaller integer than the one in the 1900th position. In other words, by choosing the critical value to be the 1900th largest point, the size of the test will be at least .05, but could be greater. There are two possible solutions to choosing a critical value for S_B.

The first option is to choose an α-level, say .05, sort the simulated values of S_B in increasing order and look in the upper tail for the area that is closest to and smaller than or equal to $\alpha = .05$ which is defined by a jump in the value of S_B. For example, when sorting the first control-control data set, the 1900th position is the integer 10. So are several positions before the 1900th which means that by choosing 10 as the critical value, the size of the test is really larger than .05. But at the 1908th position the value of S_B jumps to 11. If we use 11 as the critical value then, we have a conservative test for $\alpha = .05$.

The second option is to randomize the test, using random numbers in the decision making process. For example, in the above example, there were 23 10's before the "cutoff value" and 8 after. So, in a test situation, if the observed test statistic were 10, then the null hypothesis should be rejected with probability .26 and not rejected with probability .74. To decide whether to reject or not, use a uniform random number generator once to generate a number between 0 and 1. If the number is less than or equal to .26 then reject the null hypothesis. Otherwise accept it. Randomized tests are useful in theoretical studies, but rarely used in practice.

For this article we will use the first option discussed. For the simulated power study we will use the exact observed size for S_B and compare to tests for S_T and S_W of the same size.

Finding the critical values for S_T and S_W. For S_T and S_W, the critical value is simply the point in the simulation that represents the beginning of the upper $100 \times \alpha\%$ of the distribution. That is, with a simulation of length 2000 and $\alpha = .05$, the critical value is the 1900th largest point.

3.5 Numerical Results Comparing Test Statistics

In this section we use three different data sets to illustrate the use of the three different statistics. We also compare and contrast the simulated null distributions for each test statistic based on bootstrapping from the three different data sets. Each of the data sets consists of 20 pairs of rats observed for 15 minutes as described in Section 3.2. Two of the experiments used here are Control-Control experiments, the third is an Exposed-Control experiment. We use $\alpha = .05$ for all tests.

- CC-1: The first control-control data comes from the Forsyth Research Institute. For this experiment, no treatment whatsoever was given to any of the forty rats. The observations took place between 10 am and 2 pm and were done under white light, making the rats think that it was daytime. Rats tend to be more lethargic during the daytime hours. The results of this data set were reported in Kernan, Mullenix and Hopper (1989).

- CC-2: The second control-control data set was taken at the Iowa State University Veterinary Diagnostic Laboratory. Again, there was no treatment of any

of the rats. The observation was done between 10 am and 2 pm, but for several weeks prior to observation the rats had their time frame gradually adjusted until the observation time was "night" to the rats. These rats were observed under a red light, which made them continue to think it was night. Rats are more active at night. The results of this data set were reported in Hopper, Kernan and Wright (1990).

- EC-1: The exposed-control data set was taken from an experiment run at the Forsyth Research Institute. For this experiment the treatment was an injection of 1 mg/Kg of d-amphetamine, a drug that is commonly believed to cause behavioral differences. The control rats were injected with a similar amount of saline solution. At this dosage level, it was believed that the d-amphetamine would produce hyperactivity. The results of this data set were reported in Mullenix, Kernan, Tassinari, and Schunior (1989).

3.5.1 Test results using S_B

The calculated value of S_B for the CC-1 data set was $S_B = 1$ which is much less than the simulated critical value of 11, indicating that we should not reject the null hypothesis of no difference between the two groups of rats. Similar results were found in the CC-2 data set where $S_B = 0$ and the critical value from the simulated distribution is 11. For the EC-1 experiment, $S_B = 22$ and the critical value is 11 which leads us to reject the null hypothesis and conclude that a behavioral difference does exist. These three conclusions are, of course, what we expected from the three data sets in question, knowing that there was no treatment in the first two and that the treatment in the third data set is expected to produce a difference.

We now compare the simulated distributions of S_B across the three data sets. The difference among these data sets is manifested in the covariance structure of the test statistics, but how great a difference that would translate to for the distribution of S_B was not completely clear at the outset. Figures 3.2 (a), (b), and (c) are the histograms of the three simulated distributions and show that while there are some small differences between the three data sets, the ranges and shapes of the distributions are all about the same. We do not compute the asymptotic distributions for S_B because this distribution requires too much computation for 44-46 statistics.

3.5.2 Test results using S_T

The calculated value for S_T for the CC-1 experiment is $S_T = 35.78$. When this is compared to the critical value of 126.07 computed from the simulated distribution, the conclusion is to not reject the null hypothesis of no behavioral difference between the two groups of rats. Similar results are found in the CC-2 data set where $S_T = 23.87$ and the critical value is 113.36. For the EC-1 experiment $S_T = 284.06$ and the critical value is 111.37, strong evidence for a difference between the two groups. Again, the results agree with the expected results and with the results of S_B.

Figures 3.3 (a), (b), and (c) compare histograms of the simulated distributions of S_T from the three data sets. As we saw for S_B, the general shape and range of the distributions does not change much from one data set to the next indicating similarities over the range of covariance matrices from the three data sets. However, we do see a bit more difference here than we did for S_B. As another diagnostic, we plotted the quantiles of each pair of data sets. Figure 3.4 gives the qq-plot for CC-1 and CC-2. The straight line represents the model for which both distributions would

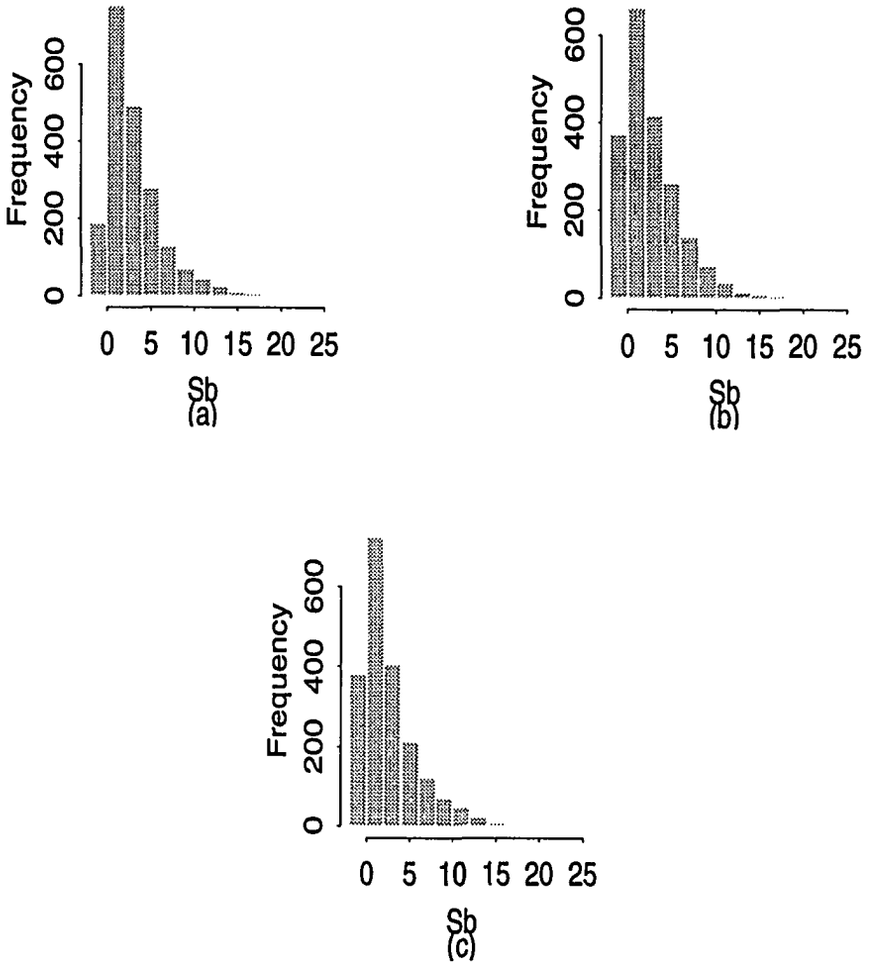

Figure 3.2: Histograms of the simulated distributions of the statistic S_B, based on data sets (a) CC-1, (b) CC-2, and (c) EC-1.

be the same. In this plot, we see that in the lower tail, the two simulated distributions are similar. In the upper tail, however, the distributions are somewhat different. This indicates that we could not substitute one distribution for the other when computing the upper quantiles needed to determine critical values for the hypothesis test. The results are similar when comparing the EC-1 data set to either the CC-1 or CC-2 data set, although the differences between the two distributions in the upper tail are more pronounced. Figure 3.5 gives the qq-plot comparing the simulated distributions for the EC-1 data set and the CC-2 data set. Here, the distributions are somewhat different as the points fall in a straight line slightly different from the theoretical line showing equality and the upper tails are very different. This means that the distribution of S_T seems to be sensitive to the type of experiment conducted, leading to the conclusion that any test based on S_T should use a critical value computed from a simulated distribution based on the same experiment.

3.5.3 Test results using S_W

For the CC-1 data set $S_W = 22.89$ compared with a critical value of 32.41. As with the other two statistics for this data set, we fail to reject the null hypothesis indicating that there is no detectable systematic behavioral difference between the two groups of rats. Again, the results are similar for the CC-2 data set which has $S_W = 22.76$ and a critical value of 33.47. The EC-1 data set does show a significant difference, again, as with the other two statistics, with $S_W = 82.95$ compared with a critical value of 35.43.

Even though these simulated distributions rely on more than the positive eigenvalues (they also depend on the corresponding normalized eigenvectors), the distri-

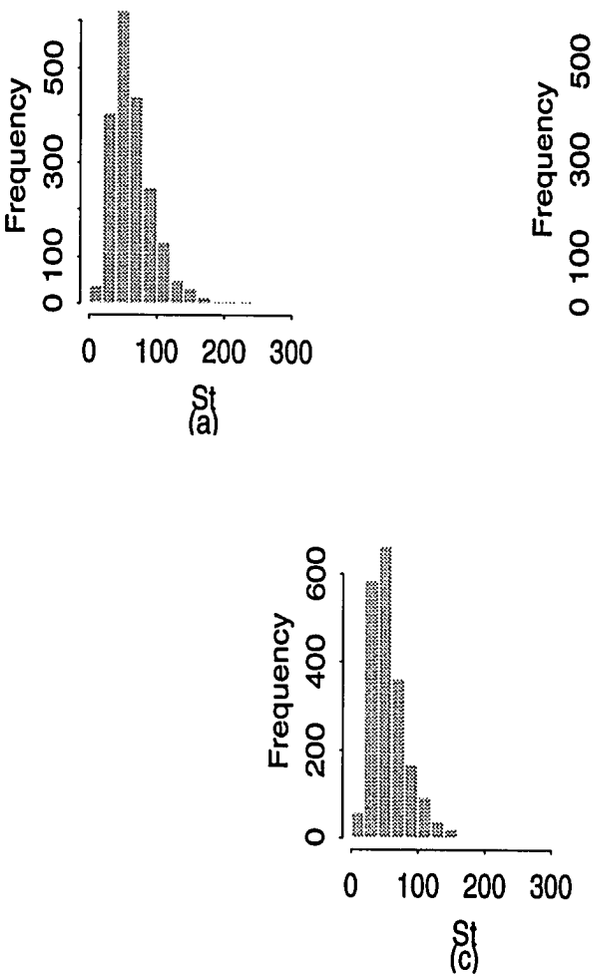

Figure 3.3: Histograms of the simulated distributions of the statistic S_T, based on data sets (a) CC-1, (b) CC-2, and (c) EC-1.

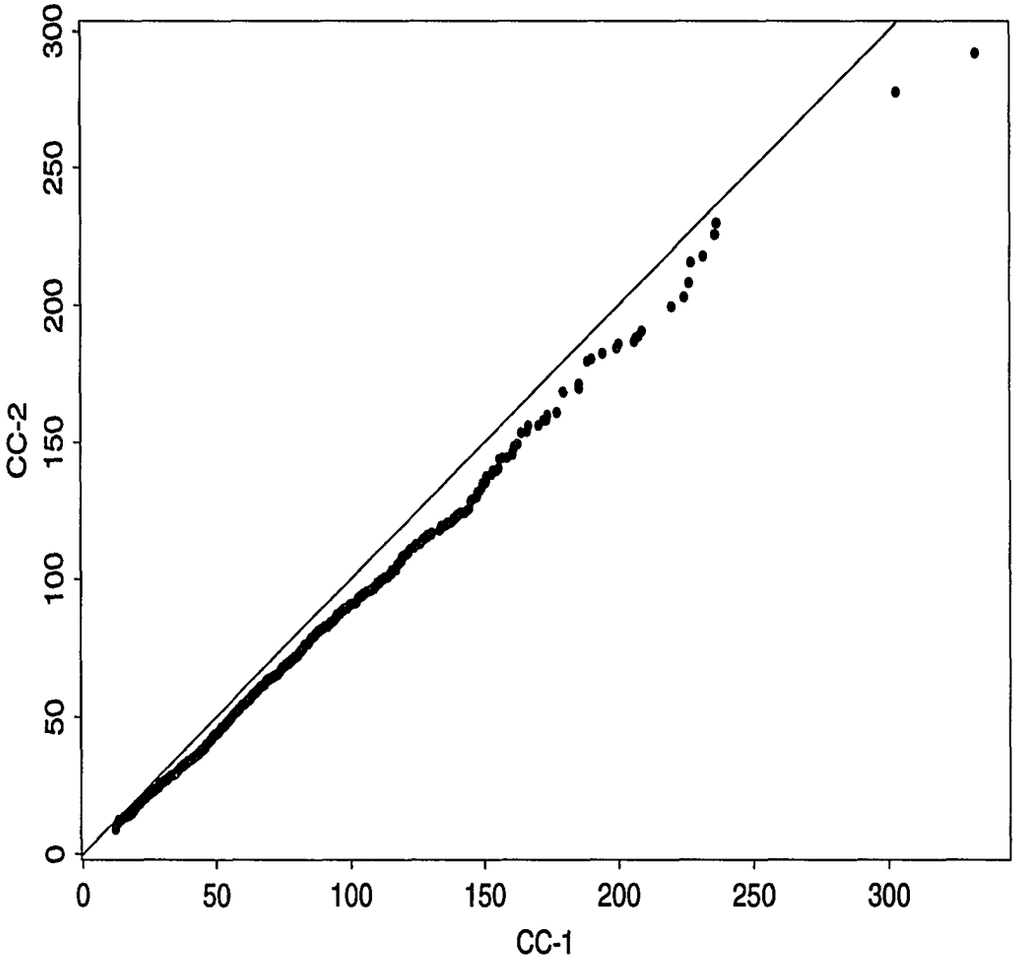

Figure 3.4: QQ-plot of the 2000 values of S_T for the two data sets CC-1 and CC-2.

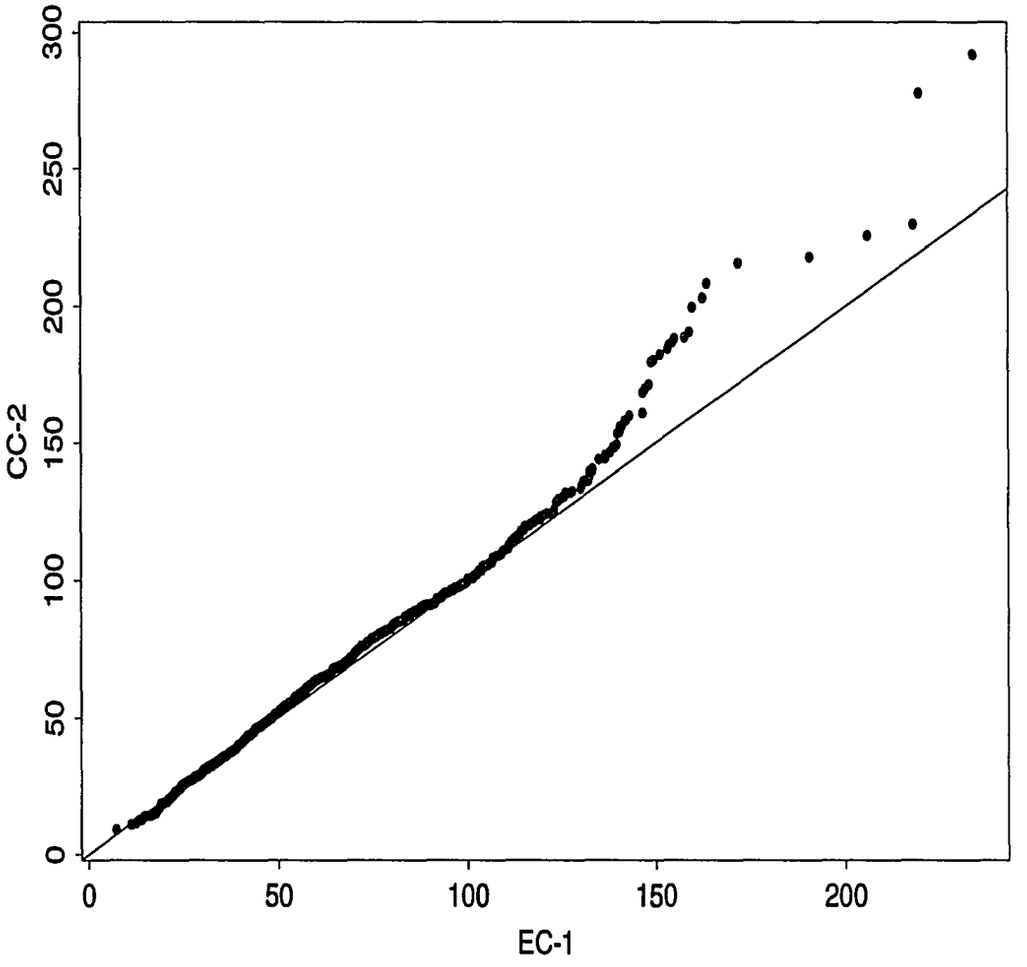

Figure 3.5: QQ-plot of the 2000 values of S_T for the two data sets EC-1 and CC-2.

butions remain basically constant from one data set to another. Figure 3.6 shows the qq-plot comparing the CC-1 and CC-2 data sets. While the points do not follow exactly the straight line indicating that the distributions are the same, the points are not far from this line. The few outliers in the far right hand tail, which represent about the top 1% of the distributions, are most likely due to Monte-Carlo sampling variability. These results are encouraging because they indicate that the statistics based on the estimated covariance matrices from both control-control experiments have approximately the same distribution. This is expected because, as we show in Section 3.6.3, the asymptotic distribution for both data sets is the same.

3.6 Asymptotic Distributions

Throughout this section, as before, $\mathbf{\Sigma}$ signifies the true covariance matrix of the test statistics. Denote the eigenvalues of $\mathbf{\Sigma}$ by λ_i and let $\underline{\mu}$ will represent the true mean vector of the test statistics, the null hypothesis being that $\underline{\mu} = (0, ..., 0)'$.

3.6.1 The asymptotic null distribution of S_B

The statistic, S_B, is a count of the number of significant paired hypothesis tests. In other words it is the sum of m Bernoulli random variables with the probability of success equaling the α level chosen for use in the original hypothesis tests. In the case where all of the tests are independent of one another, S_B would simply be a binomial random variable with parameters m and α. In the case where all of the statistics are completely correlated, that is all correlations are equal to 1,

$$S_B = \begin{cases} 0 & \text{with probability } 1-\alpha, \\ m & \text{with probability } \alpha \end{cases}$$

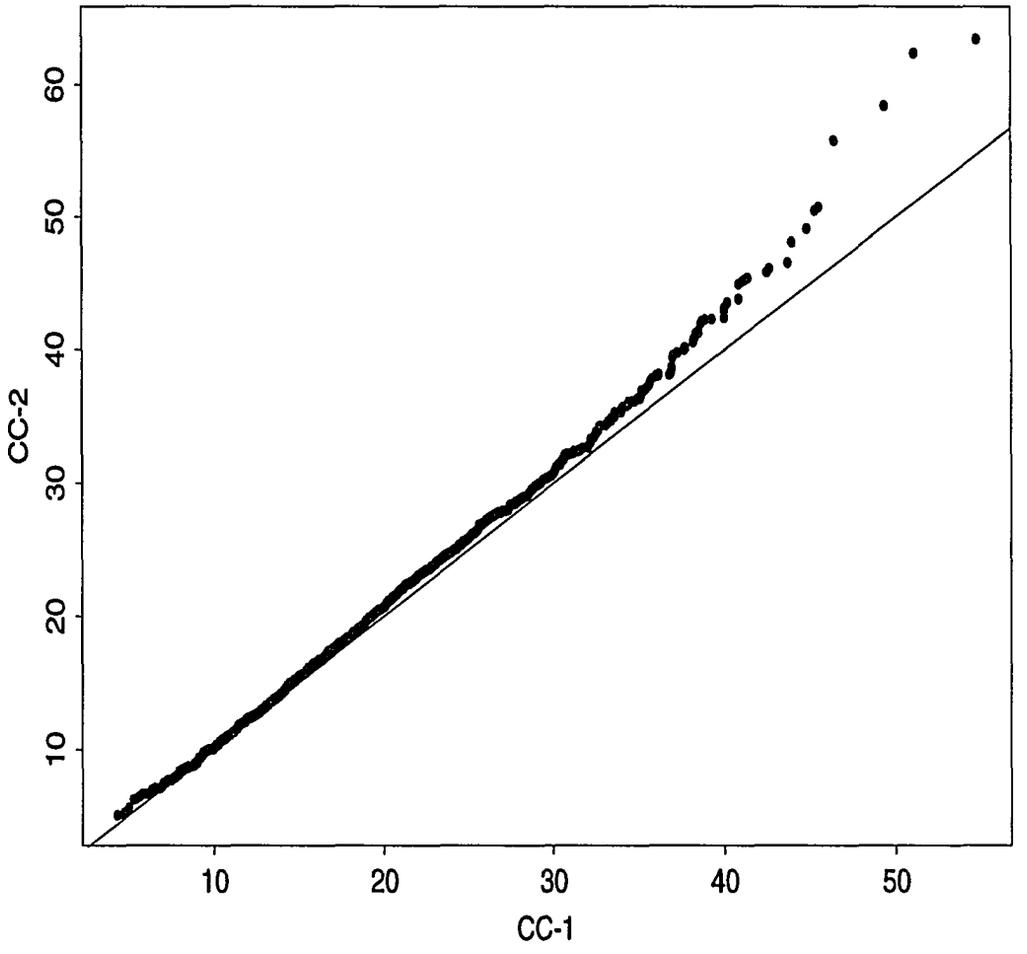

Figure 3.6: QQ-plot of the 2000 values of S_W for the two data sets CC-1 and CC-2.

so $\frac{1}{m}S_B$ is a Bernoulli random variable with probability of success α. For all other cases, the probability distribution is more complicated.

Let $\underline{a} = (a_1, ..., a_m)$ be the vector of decisions for the tests. That is, $a_i = 0$ if the ith test was not significant and $a_i = 1$ otherwise. Then the discrete multivariate probability density function for the vector of the m decisions can be written as (Bahadur, 1961):

$$p(\underline{a}) = \alpha^{\sum_{i=1}^{m} a_i}(1-\alpha)^{m-\sum_{i=1}^{m} a_i} f(\underline{a})$$

where

$$f(\underline{a}) = 1 + \sum_{i<j} r_{ij} z_i z_j + \sum_{i<j<k} r_{ijk} z_i z_j z_k + ... + r_{12...n} z_1 \cdots z_n$$

with

$$z_i = \frac{a_i - \alpha}{[\alpha(1-\alpha)]^{1/2}}$$

and

$$\begin{aligned} r_{ij} &= \mathrm{E}(z_i z_j) \\ r_{ijk} &= \mathrm{E}(z_i z_j z_k) \\ &... \\ &... \\ r_{12...n} &= \mathrm{E}(z_1 \cdots z_n). \end{aligned}$$

Now, $S_B = \sum_{i=1}^{m} a_i$, is the statistic of interest. To develop the discrete probability density function for S_B, let A_b be the set of decision vectors which have exactly b decisions rejecting the null hypothesis. That is $A_b = \{\underline{a} : \sum_{i=1}^{m} a_i = b\}$. Then

(Bahadur, 1961)

$$\begin{aligned} \Pr(S_B = b) &= \sum_{\underline{a} \epsilon A_b} p(\underline{a}) \\ &= \alpha^b (1-\alpha)^{m-b} \sum_{\underline{a} \epsilon A_b} f(\underline{a}). \end{aligned} \tag{3.5}$$

3.6.2 The asymptotic power of S_B

To calculate the asymptotic power of S_B one must first calculate the critical value of the statistic for a given α under the null hypothesis. Denote that critical value by b_C and let α_i be the true probability that $a_i = 1$ under the alternative hypothesis. Then the asymptotic power of this statistic can be written as (Bahadur, 1961):

$$\text{power}_{S_B} = \Pr\left(S_b \geq b_C\right) = \sum_{\underline{a} \epsilon A_{b_C}} \left(\Pi \alpha_i^{a_i} (1-\alpha_i)^{1-a_i}\right) f(\underline{a}) \tag{3.6}$$

where b_C is the critical value and $A_{b_C} = \{\underline{a} : \sum_{i=1}^{m} a_i = b_C\}$. Because these calculations become prohibitively long as more test statistics become involved, we rely on simulation results in the Section 3.7 to compare the simulated power of the three statistics.

3.6.3 The asymptotic null distribution of S_T

Now consider the second statistic, S_T, the sum of the individual test statistics squared. Each of the paired tests for the number of initiations and total time was based on a sample of n pairs, leading to a number of t-statistics, each approximately following a Student's t Distribution with $n-1$ degrees of freedom. The tests based on the K-functions lead to test statistics that are approximately normally distributed.

We assume that, under the null hypothesis, when there is no difference between the two groups, each paired t test will lead to a central t-statistic and each K function test will lead to a standard normal statistic. Therefore, asymptotically (with large n), each statistic has a standard normal distribution. This in turn means that each t^2 statistic, asymptotically, has a χ^2_1 distribution. S_T, then, asymptotically, has the distribution of the sum of correlated χ^2_1 random variables. Gordon and Ramig (1983) show that the distribution function of the sum of correlated χ^2_n random variables can be evaluated as follows:

$$\begin{aligned}\Pr(x_1 \le S_T \le x_2) &= F(x_2) - F(x_1) \\ &= \frac{\cos(n\pi)}{\pi} \lim_{T\to\infty} \int_0^T \frac{r_1(t) r_2(t) \sin\left[\phi_1(t) - \frac{n\theta(t)}{2}\right] dt}{t} \qquad (3.7)\end{aligned}$$

where r_1, r_2, ϕ_1 and θ are defined below:

$$\begin{aligned} r_1(t) &= \left[A^2(t) + B^2(t)\right]^{1/2} \\ A(t) &= \cos(x_1 t) - \cos(x_2 t) \\ B(t) &= \sin(x_2 t) - \sin(x_1 t) \\ r_2(t) &= \{\left[a^2(t) + b^2(t)\right]^{1/2}\}^{-n/2} \\ a(t) &= 1 - 4t^2 \sum_{j_1<j_2} (\lambda_{j_1}\lambda_{j_2}) + 16t^4 \sum_{j_1<j_2<j_3<j_4} (\lambda_{j_1}\lambda_{j_2}\lambda_{j_3}\lambda_{j_4}) \\ &\quad + \ldots + (-1)^k (2t)^{2k} \sum_{j_1<j_2<\ldots<j_{2k}} ((\lambda_{j_1}\lambda_{j_2}\cdot\ldots\cdot\lambda_{j_{2k}}) \end{aligned}$$

and if m (the number of statistics being added) is even:

$$\begin{aligned} b(t) &= -2t \sum_{j=1}^{m} \lambda_j + 8t^3 \sum_{j_1<j_2<j_3} (\lambda_{j_1}\lambda_{j_2}\lambda_{j_3}) \\ &\quad + \ldots + (-1)^{k+1} (2t)^{2k-1} \sum_{j_1<j_2<\ldots<j_{2k-1}} (\lambda_{j_1}\lambda_{j_2}\cdot\ldots\cdot\lambda_{j_{2k-1}}) \end{aligned}$$

where the λ_j's are the eigenvalues of the covariance matrix $\mathbf{\Sigma}$. If m is odd the last term of $b(t)$ is

$$(-1)^{k+1}(2t)^{2k+1} \sum_{j_1<j_2<\ldots<j_{2k+1}} (\lambda_{j_1}\lambda_{j_2}\cdot\ldots\cdot\lambda_{j_{2k+1}}).$$

In this definition $\phi_1(t)$ is the angle that lies in the quadrant in which $(A(t), B(t))$ lies, and has, as its angle of reference, $\arctan|B(t)/A(t)|$. In other words, $\phi_1(t)$ is defined as follows: let $\arctan|B(t)/A(t)| = \gamma$ and i be the quadrant in which $(A(t), B(t))$ lies. Then

$$\phi_1(t) = (-1)^{i+1}\gamma + \left|\left|\frac{i}{2}\right|\right|\pi$$

where $||i/2||$ signifies the greatest integer less than or equal to $i/2$. Similarly $\theta(t)$ lies in the quadrant in which $(a(t), b(t))$ lies and has, as its angle of reference, $\arctan|b(t)/a(t)|$. It should be noted that this distribution depends on $\mathbf{\Sigma}$ only through the eigenvalues of $\mathbf{\Sigma}$, so that two experiments with different values for $\mathbf{\Sigma}$, but similar eigenvalue structures will have approximately the same asymptotic distribution.

We now compare the simulated distributions for each of the data sets from Section 3.5 with the asymptotic distribution based on the same covariance matrix. Because the asymptotic distribution depends on the positive eigenvalues of the covariance matrix, as in Section 3.4.3.2, we assume that there are 19, 19, and 20 positive eigenvalues respectively for data sets CC-1, CC-2, and EC-1.

Figures 3.7 (a), (b), and (c) compare the asymptotic null distribution CDF computed using eq. (3.6) with the null distribution CDF computed using the simulated values of S_T for each of the data sets. In all these plots the functions nearly overlap suggesting that they are very similar. Therefore, for S_T, critical values for hypothesis tests could be computed using either the asymptotic null distribution or the simulated

null distribution.

3.6.4 The asymptotic power of S_T

We do not present a closed form of the asymptotic power of S_T. The non-null asymptotic distribution of S_T is that of the sum of correlated non-central χ^2 variables which does not have a closed form to our knowledge. It would, however, be possible to simulate the asymptotic power.

3.6.5 The asymptotic distribution of S_W

The Wald-like statistic, S_W, has, as its asymptotic distribution, a noncentral χ^2_r distribution with noncentrality parameter $(1/2)\underline{\mu}'\mathbf{A}\underline{\mu}$, where r is the rank of $\mathbf{\Sigma}$ and $\mathbf{A} = \mathbf{Q}_1\Lambda_1^{-1}\mathbf{Q}_1'$ is computed from the eigenvalues and normalized eigenvectors of $\mathbf{\Sigma}$. The proof of this result can be found in Appendix B. Note that the null distribution (the central χ^2_r distribution) does not depend on $\mathbf{\Sigma}$, or even its specific eigenvalue structure, but instead, only on the rank of $\mathbf{\Sigma}$.

Figure 3.8 gives both the asymptotic distribution (a χ^2_{19}) and the two simulated distributions for the CC-1 and CC-2 data sets. When looking at Figure 3.8, it appears that the simulated CDFs are not quite as close to the asymptotic CDF as for S_T, but are still quite similar. It does appear that the asymptotic CDF increases slightly faster than the simulated CDFs indicating that using critical values computed from the asymptotic distribution would be conservative and could therefore be used in a hypothesis test. It is also clear from Figure 3.8 that the two simulated distributions are very similar. The results above also hold when comparing the EC-1 CDF to the corresponding asymptotic CDF (a χ^2_{20}).

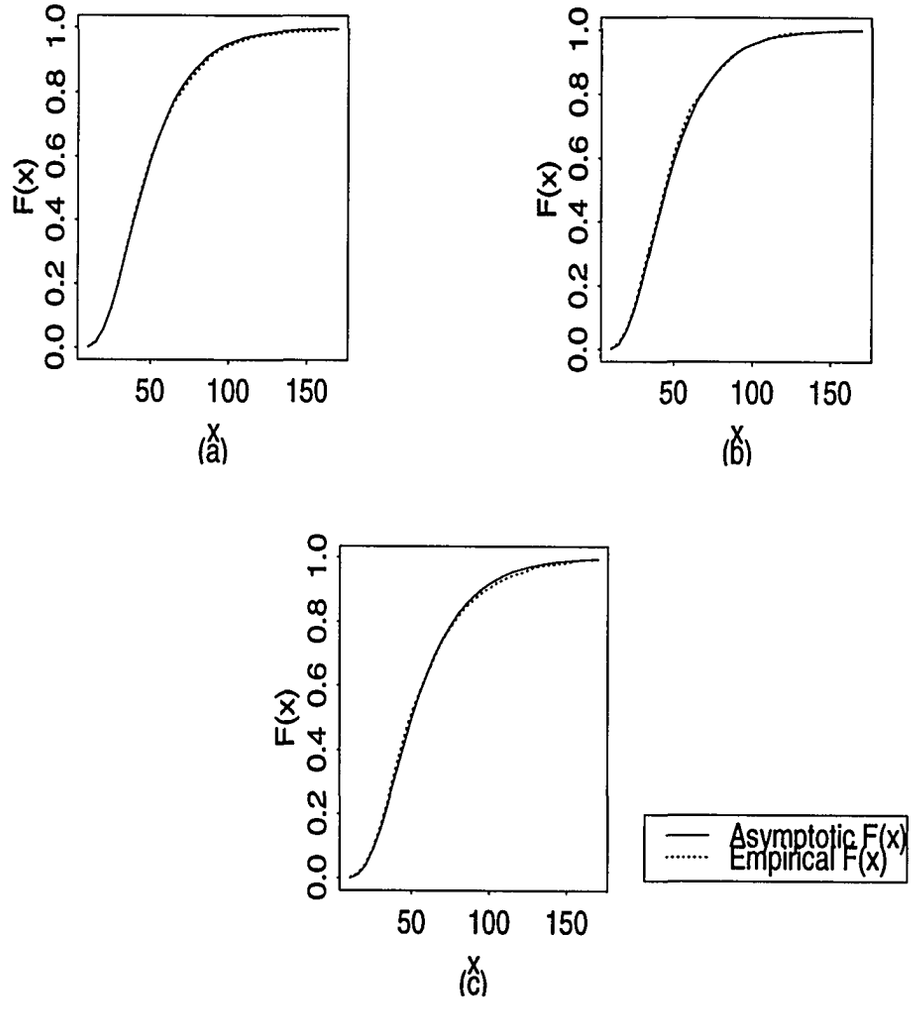

Figure 3.7: Asymptotic CDF and Simulated CDF for the three data sets: (a) CC-1, (b) CC-2, and (c) EC-1.

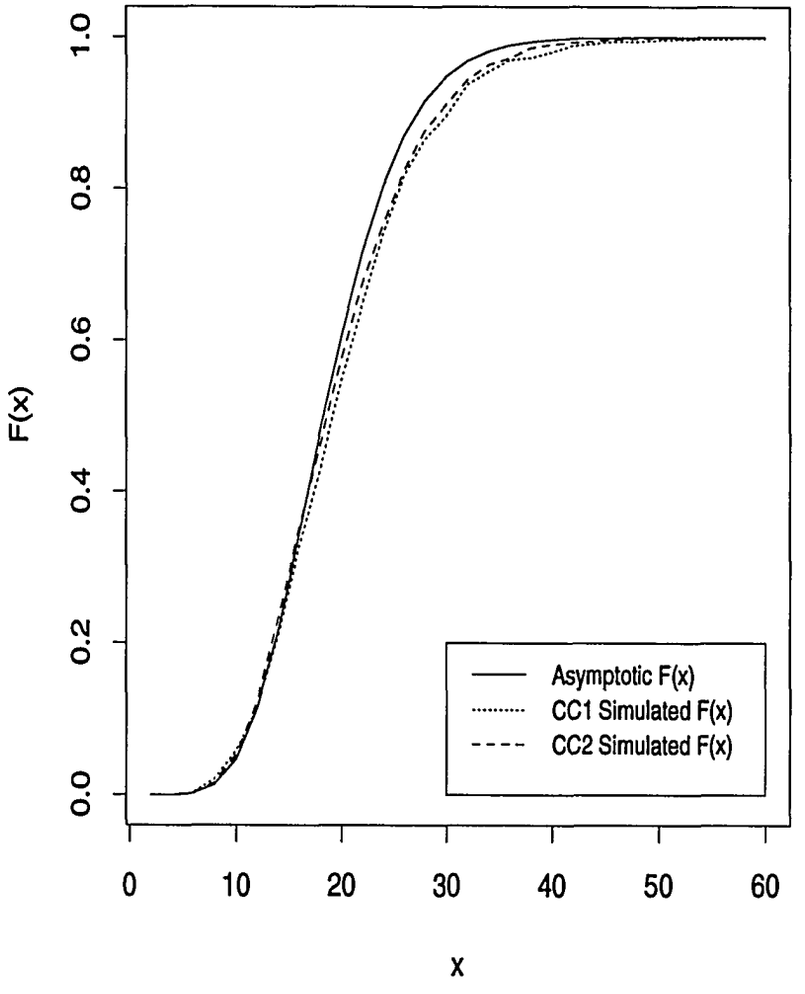

Figure 3.8: Asymptotic CDF (a χ^2_{19}) and the simulated CDF from CC-1 and CC-2

3.6.6 The asymptotic power of S_W

The asymptotic power of S_W can be approximated from the asymptotic distribution of S_W as follows:

$$\begin{aligned} \text{Power}_{S_W} &= \Pr\left(S_W > x(r,\alpha)\right) \\ &\cong \Pr\left[\chi^2_r\left(\frac{1}{2}\underline{\mu}'\mathbf{A}\underline{\mu}\right) > x(r,\alpha)\right], \end{aligned}$$

where $x(r,\alpha)$ is the appropriate critical value for a central χ^2_r random variable. Therefore, letting $F(x)$ be the CDF of a $\chi^2_r\left(\frac{1}{2}\underline{\mu}'\mathbf{A}\underline{\mu}\right)$ random variable and letting $\delta = \underline{\mu}'\mathbf{A}\underline{\mu}$ be the noncentrality parameter (Johnson and Kotz, 1970)

$$\text{Power}_{S_W} = \Pr[\chi^2(\frac{1}{2}\delta) > x(r,\alpha)]$$

$$= 1 - e^{\frac{1}{2}\delta} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\delta)^j}{j!} \frac{1}{2^{(1/2)r+j}\Gamma(\frac{1}{2}r+j)} \int_0^{x(r,\alpha)} y^{(1/2)r+j-1} e^{-(1/2)y} dy.$$

3.7 Power Study to Compare the 3 Statistics

3.7.1 Computation of simulated power

For this simulated power study, we limit ourselves to a sample size of 20 pairs (the typical sample size for these experiments) and several values of α chosen specifically for each data set dependent on the distribution of S_B, as described in Section 3.4.3.4. We also limit the alternative distributions to being the same as the null distributions but with a shifted mean. In fact, we look at the case where each element of $\underline{\mu}$ is shifted by the same standardized increment. That is, we look at the case where $\underline{\mu} = \mu\underline{1}$, where μ is, for example .25 instead of the value 0 from the null hypothesis.

Figures 3.9, 3.10, and 3.11 show the simulated power curves for the three models, each using a covariance matrix estimated in one of the three data sets, as the value of μ increases. Note that each model has a different value of α because of the simulated distribution of S_B, as described above, but we chose in each case to use the value of α that was closest to, but larger than .05. The value of the simulated power is computed though simulation for each of the three statistics at the nine points where the test statistics have means of 0, .25, .5, ..., 2.00, respectively. The curves have then been smoothed using a spline function. Technically, the simulated power could have been computed at more points so that a spline would not be necessary, but the simulated power of S_W at each point took about 60 cpu hours to compute.

3.7.2 Comparison of simulated power

Initially we had two expectations about the comparison of the three statistics, both of which were realized, at least to some extent. The graphs comparing the three statistics for each of the three models used are in Figures 3.9 - 3.11. The expectations and rationales were:

- S_T is more powerful than S_B. S_T carries with it information about the strength of the significance which S_B does not have.

- S_W is much more powerful than either S_T or S_B. S_W uses the information about the correlational structure between the individual test statistics which neither S_B or S_T do. And, like S_T, S_W carries the information about the strength of the significance of the individual tests.

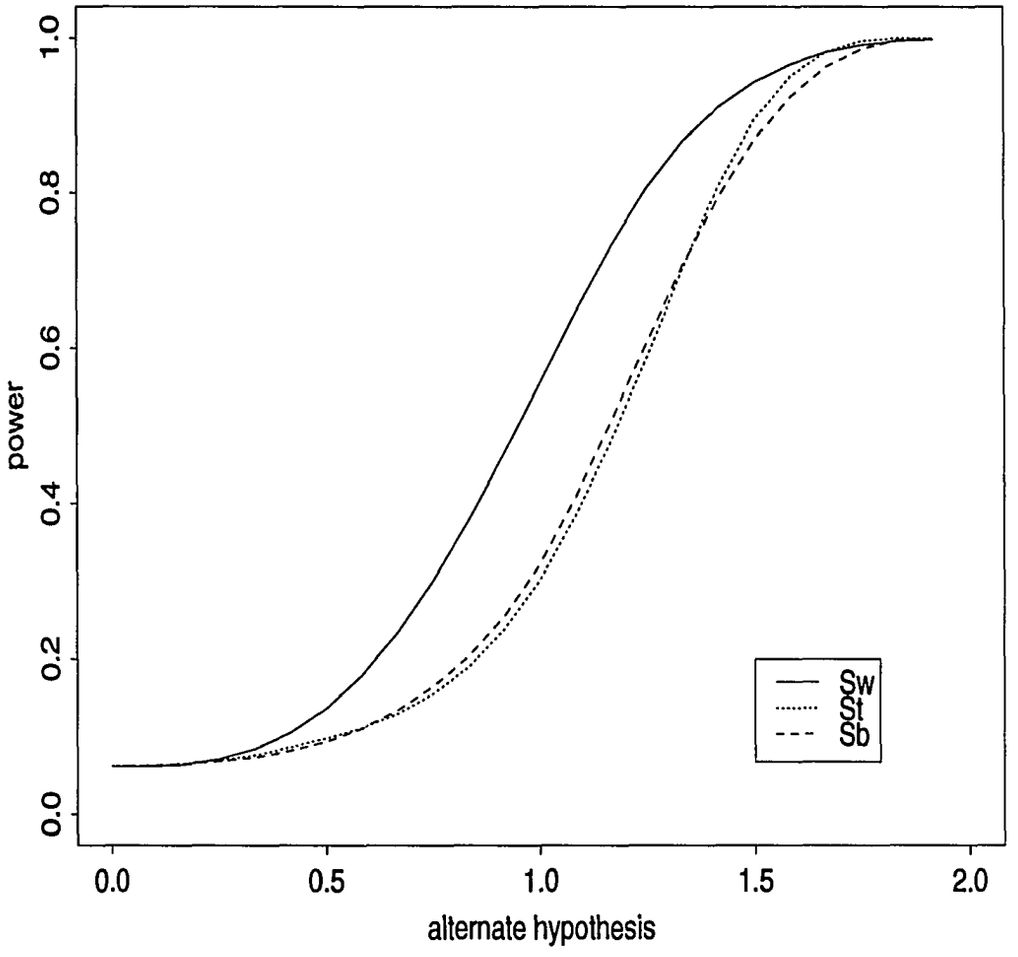

Figure 3.9: Simulated power curve for the model using the covariance matrix estimated from the CC-1 data set

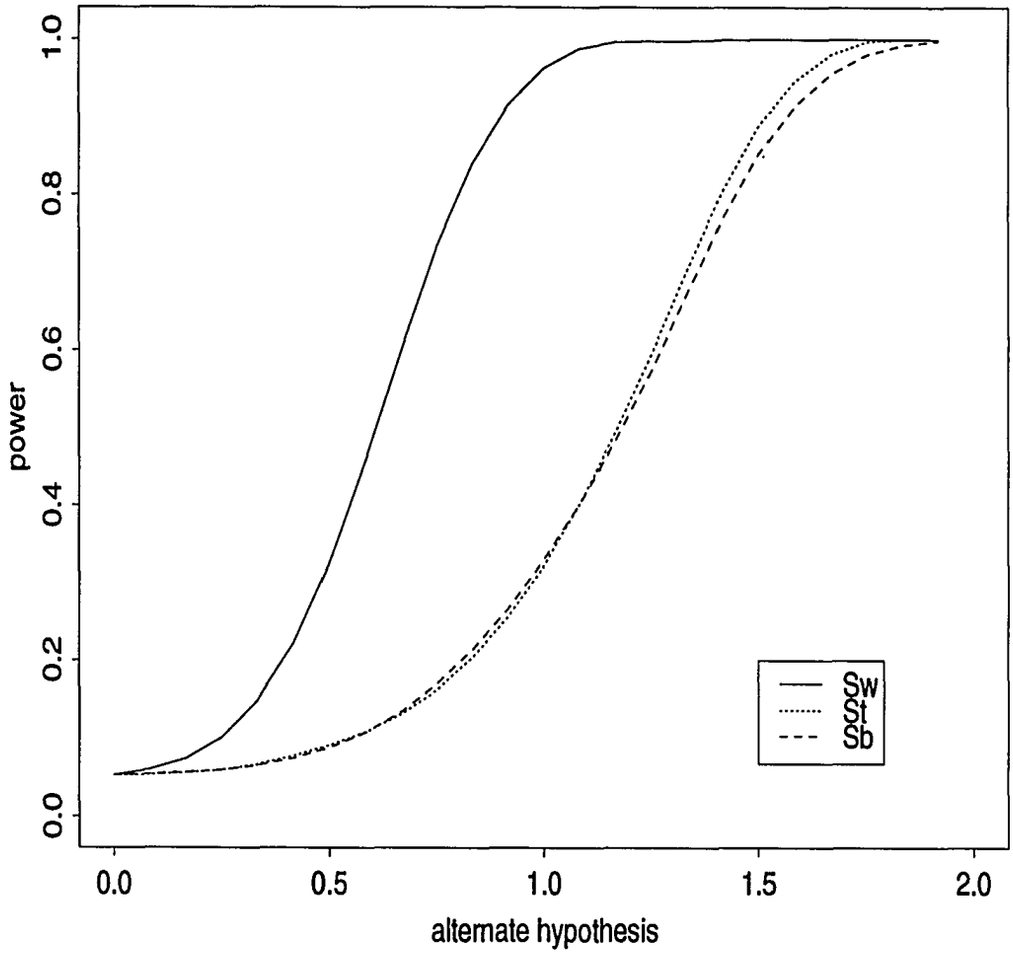

Figure 3.10: Simulated power curve for the model using the covariance matrix estimated from the CC-2 data set

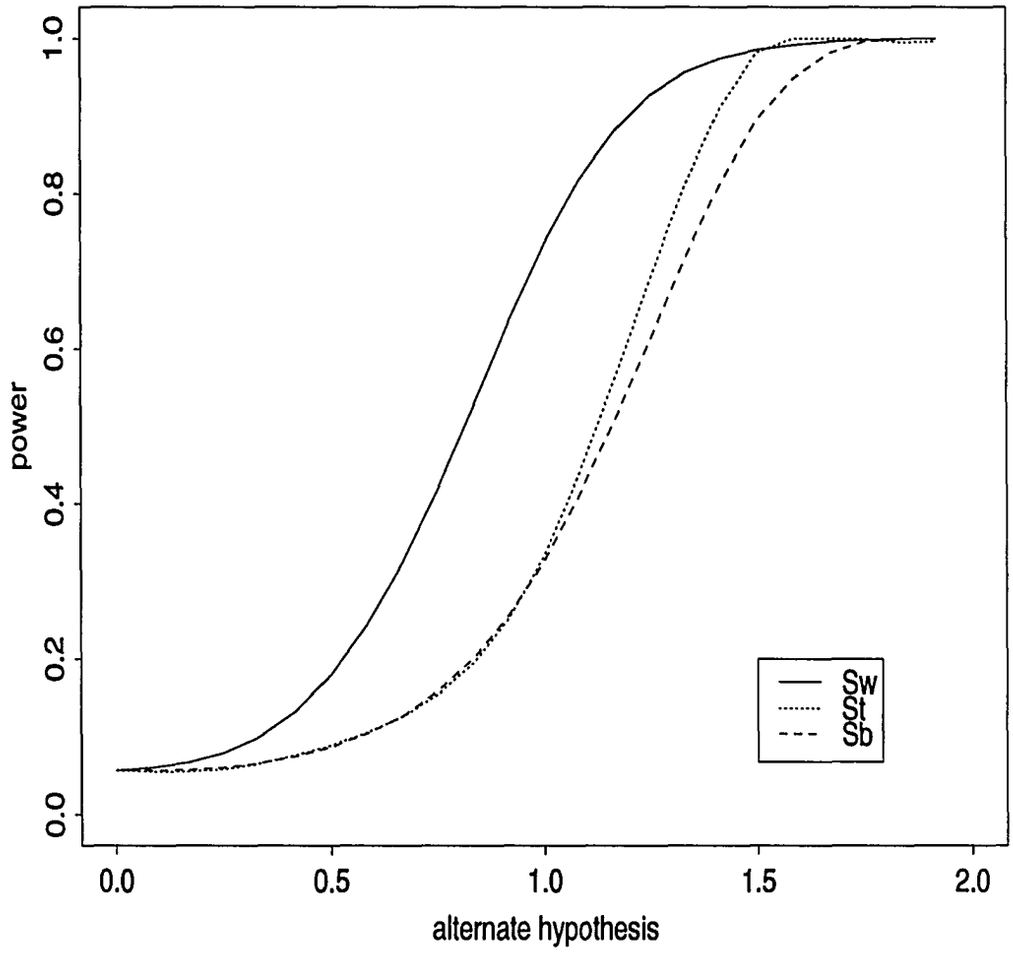

Figure 3.11: Simulated power curve for the model using the covariance matrix estimated from the EC-1 data set

Because the power for this study was computed through simulation, an estimate of the Monte Carlo error was computed to check the variability of the simulated power calculations. To compute the Monte Carlo error, one point on one simulated power curve for S_W was chosen and replicated 8 times. We chose to use S_W since it should have the most error associated with its simulation due to the sub-bootstrapping. These 8 values of power had a mean of .84 and a standard deviation of 0.0055, indicating that the error due to simulation was quite small.

3.7.3 Conclusions from the simulated power study

From the three data sets used in this article, we have shown that S_T does about the same as (though usually slightly better than) S_B, as expected. Also, S_W performs much better than either S_T or S_B under all three proposed models. This suggests that S_W is superior to the other two statistics and should be the statistic used for analysis of this kind.

3.8 Discussion and Areas for Future Research

In this article we have defined and compared three test statistics designed to combine the information from many dependent hypothesis test statistics into one test statistic. These three statistics are: S_B, a count of the number of significant individual tests; S_T, the sum of the squared individual test statistics; and S_W, a Wald-like combination of the individual test statistics, making use of the statistics' covariance matrix. S_B has been in use in this type of analysis for several years (e.g., Kernan and Meeker, 1992).

We have shown through simulation studies using covariance matrices based on

three real data sets, that the simulated power of S_W is much higher than that of either S_B or S_T. We also found that S_T is generally slightly better than S_B, though the difference is not great. These results lead us to suggest using S_W.

The problem that we attempted to address by creating one test statistic out of a set of hypothesis tests, was the problem of multiplicity of hypothesis tests. If 20 tests are performed at $\alpha = .05$ and if all null hypotheses are true, we would "expect" to see one false positive. In other words, the size of the combination of the tests will be much larger than the size of the individual tests.

There are several other methods available for dealing with multiple hypothesis tests. Perhaps the simplest to use is Bonferroni's method which adjusts the α-levels used for the individual tests to α_0/k, where k is the number of tests and α_0 is a conservative estimate of the overall α-level of the combined results. This method gives results based on a conservative α-level and does very poorly for large k. Each of our three tests does better than Bonferroni because we have large k and are able to use an (approximately) exact α-level. Also, the Bonferroni method does not take the correlational structure between the test statistics into consideration, whereas S_W does.

The work done for this article has brought to light several areas for future research. One question that we have yet to answer is how the length of the observation period would change the power of the statistics. Most likely a shorter period of observation time would not be preferred because more variability would be introduced into the data. And, 15 minutes was chosen because, in practice, once rats become used to an environment they reduce their level of activity so much that little more information could be gained. But, perhaps after 15 minutes they could be moved to

another novel environment and observed for a longer time period.

One problem that arose in this research is that there was much rat-to-rat variability. In order to reduce this variability somewhat, we suggest, when possible, observing both the controls and the treated rats before administering the treatment and then observing them all again after the treatment. This is not always possible because frequently the treatment is given to pregnant rats and the offspring are observed. Also, the question of the novel environment occurs. If the rats are put into this environment as controls and then again as treated rats, the second time the environment is no longer novel. For this type of design, a second novel environment must be constructed that is similar enough not to give an environment effect, but different enough to be novel to the rats.

A third question is which test statistics to include in the overall statistic. In this article we have addressed combining information from several applications each of three types of statistics. Kernan and Meeker (1992) used four types in their analysis. We have not done a comprehensive study on which individual statistics to include. There may be unnecessary redundancy or there may be other statistics, not yet considered, that should be included.

3.9 Acknowledgments

The authors wish to thank Dr. William J. Kernan of the Veterinary Diagnostic Laboratory, Iowa State University, both for supplying the data used in this study and for his continuing support of the project. Several times he made himself available for discussions pertaining to this problem. Also, Dr. Yasuo Amemiya and Dr. Krishna Athreya, of the Department of Statistics, Iowa State University, offered several very

useful suggestions which have been incorporated into this document. Finally, we wish to express our gratitude to Kathy Shelley and Bud Meador in the Department of Statistics, Iowa State University, and John Hascall of the Iowa State University Computation Center for their support and help through the difficulties encountered during the 6000 or so cpu hours of computing required for this project.

3.10 Bibliography

Amemiya, Y., and Fuller, W.A. (1984) "Estimation for the multivariate errors-in-variables model with estimated error covariance matrix." *The Annals of Statistics*, 12, 497-509.

Bahadur, R.R. (1961) A representation of the joint distribution of responses to n dichotomous items. In *Studies in Item Analysis and Prediction*, H. Solomon (ed.), Stanford, California: Stanford University Press.

Baumeister, A.A. (1978) Origins and Control of Stereotyped Movements. Quality of Life in Severely and Profoundly Mentally Retarded People: Research Foundations for Improvement. Washington D.C., American Association in Mental Deficiency, Monograph #3, 353-384.

Bickel, P.J., and Freedman, D.A. (1981) "Some Asymptotic Theory for the Bootstrap." *The Annals of Statistics*, 9, 1196-1217.

Cannon, A.R., and Cressie, N.A.C. (1993) "Temporal Analogues to Spatial K Functions." *Iowa State University Preprint* #93-8.

Cattell, R.B. (1966) "The scree test for the number of factors." *Multivariate Behavioral Research*, 1, 245-276.

Chow, Y.S., and Teicher, H. (1988) *Probability Theory: Independence, Interchangeability, Martingales.* New York: Springer Verlag.

Cressie, N.A.C. (1991) *Statistics for Spatial Data.* New York: Wiley - Interscience.

Diggle, P.J. (1983) *Statistical Analysis of Spatial Point Patterns.* New York: Academic Press.

Franklin, J. N. (1968) *Matrix Theory.* Englewood Cliffs, New Jersey: Prentice Hall.

Gordon, N. H., and Ramig, P. F. (1983) "Cumulative Distribution Function of the Sum of Correlated Chi-Squared Random Variables." *Journal of Statistical Computing and Simulation*, 17, 1–9.

Haccou, P., and Meelis, E. (1992) *Statistical Analysis of Behavioural Data.* Oxford: Oxford University Press.

Hopper, D. L., Kernan, W. J., and Wright, J. R. (1990) "Computer pattern recognition: An automated method for evaluating motor activity and testing for neurotoxicity. *Neurotoxicology and Teratology*, 12, 419-428.

Johnson, L. J., and Kotz, K (1970) *Continuous Univariate Distributions-2.* Boston: Houghton Mifflin Company.

Kernan, W. J., and Meeker, W. Q. (1992) "A Statistical Test to Assess Changes in Spontaneous Behavior of Rats Observed with a Computer Pattern Recognition System." *Journal of Biopharmaceutical Statistics*, 2, 115–135.

Kernan, W. J. Jr., Mullenix, P. J., and Hopper, D. L. (1989) "Time Structure Analysis of Behavioral Acts Using a Computer Pattern Recognition System." *Pharmacology Biochemistry and Behavior*, 34, 863–869.

Kernan, W. J. Jr., Mullenix, P. J., Kent, R., Hopper, D. L., and Cressie, N. A. C. (1988) "Analysis of the Time Distribution and Time Sequence of Behavioral Acts." *International Journal of Neuroscience*, 43, 35–51.

Mullenix, P. J., Kernan, W. J., Tassinari, M. S., and Schunior, A. (1989) "Generation of dose-response data using activity measures." *Journal of the American College of Toxicology*, 8, 185-197.

Norton, S., Mullenix, P., and Culver, B. (1976) "Comparison of the Structure of Hyperactive Behavior in Rats after Brain Damage from X-irradiation, Carbon Monoxide and Pallidal Lesions." *Brain Research*, 116, 49–67.

Pohl, P. (1976) "Spontaneous Fluctuation in Rate of Body Rocking: A Methodological Note". *Journal of Mental Deficiency Research*, 20, 61-65.

Pontius, A. A. (1973) "Dysfunction Patterns Analogous to Frontal Lobe System and Caudate Nucleus Syndrome in some Groups of Minimal Brain

Dysfunction." *Journal of American Medical Women's Association*, 28, 285-291.

Searle, S. R. (1971) *Linear Models.* New York: John Wiley & Sons, Inc.

Wender, P. H. (1971) *Minimal Brain Dysfunction in Children.* New York: Wiley - Interscience.

Westfall, P. H., and Young, S. S. (1989) "p Value Adjustments for Multiple Tests in Multivariate Binomial Models." *Journal of the American Statistical Association*, 84, 780-786.

3.11 Appendix A. Convergence of a Sample Covariance Matrix

In this appendix we prove that the estimated covariance matrix of a set of test statistics computed by bootstrapping the original sample (as described in Section 3.4.3.3) will converge in conditional probability to the theoretical covariance matrix of the original test statistics. To see this, let $X_1, X_2, \ldots X_n$ be the original sample of n vectors of paired differences and let

$$\tilde{t}_{nk} = \frac{\sqrt{n}\left(\bar{X}_{nk} - \mu_k\right)}{s_{nk} \wedge \delta_n}$$

where $\bar{X}_{nk}$ is the sample mean of the kth statistic, s_{nk} is the sample standard deviation of the kth statistic, and δ_n is a sequence for which $\delta_n \to 0$ and $n\delta_n^4 \to \infty$. This statistic is slightly different than the usual definition of the t statistic because the variance and covariance of t statistics are not always defined. The use of the sequence, δ_n, assures the existence of the relevant variances and covariances. In practice $\tilde{t}_{nk}$ will be the same as the usual t statistic (notated here by t_{nk}) because δ_n can be taken to be very small. And, we will show that the limit of the covariance matrix based on the vector of $\tilde{t}_{nk}$'s is the same as the limit of the covariance matrix based on the vector of t_{nk}'s as $n \to \infty$. We will assume that $\mathrm{E}(X_{ik}^4) < \infty$ for all k.

Let $Y_1^{(j)}, Y_2^{(j)}, \ldots Y_n^{(j)}$ be the jth bootstrap sample taken from $X_1, X_2, \ldots X_n$

and define

$$\tilde{t}_{nk}^{*(j)} \equiv \frac{\sqrt{n}\left(\bar{Y}_k^{(j)} - \bar{X}_k\right)}{s_k^{*(j)} \wedge \delta_n}.$$

Letting $j = 1, \ldots N$, we compute the estimated covariance matrix from the N bootstrap samples as

$$\tilde{\Sigma}_{Nn}^{*} = \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{\mathbf{t}}_{n.}^{*(i)} - \bar{\tilde{\mathbf{t}}}_{n.}^{*(i)}\right)\left(\tilde{\mathbf{t}}_{n.}^{*(i)} - \bar{\tilde{\mathbf{t}}}_{n.}^{*(i)}\right)'.$$

We will show that $\tilde{\Sigma}_{Nn}^{*}$ is a consistent estimator for Σ, the true covariance matrix of the original test statistics. We start by finding Σ. Then we show that $\tilde{\Sigma}_{Nn}^{*}$ converges to Σ. Finally, we show that Σ_{Nn}^{*}, the estimated covariance matrix based on the usual bootstrap t-statistics, converges to Σ.

Let $\mathrm{cov}(X_{1k}, X_{1k'}) = \sigma_{kk'}$ and $\mathrm{var}(X_{1k}) = \sigma_k^2$. Now, rewrite $\tilde{t}_{nk}$ as follows:

$$\begin{aligned}\tilde{t}_{nk} &= \frac{\sqrt{n}\left(\bar{X}_{nk} - \mu_k\right)}{\sigma_k}\left(\frac{\sigma_k}{s_{nk} \wedge \delta_n} - 1\right) + \frac{\sqrt{n}\left(\bar{X}_{nk} - \mu_k\right)}{\sigma_k} \\ &= r_{nk} + p_{nk}\end{aligned}$$

and similarly $\tilde{t}_{nk'} = r_{nk'} + p_{nk'}$. The covariance of $\tilde{t}_{nk}$ and $\tilde{t}_{nk'}$ is now a function of $\mathrm{cov}(p_{nk}, p_{nk'})$, $\mathrm{cov}(r_{nk}, p_{nk'})$, $\mathrm{cov}(r_{nk'}, p_{nk})$, and $\mathrm{cov}(r_{nk}, r_{nk'})$. We start with $\mathrm{cov}(p_{nk}, p_{nk'})$.

$$\begin{aligned}\mathrm{cov}(p_{nk}, p_{nk'}) &= \frac{n}{\sigma_k \sigma_{k'}}\mathrm{E}\left(\left(\bar{X}_{nk} - \mu_{nk}\right)\left(\bar{X}_{nk'} - \mu_{nk'}\right)\right) \\ &= \frac{\sigma_{kk'}}{\sigma_k \sigma_{k'}}.\end{aligned}$$

We claim that the other three terms go to 0 as $n \to \infty$ and to prove this claim we use the following two lemmas:

Lemma 1: If $X_1, X_2, \ldots X_n$ are iid random variable such that $\mathrm{E}(X_i^4) < \infty$, and s_n is the sample standard deviation of $X_1, X_2, \ldots X_n$, then $\mathrm{E}(\sigma - s_n)^4 = O(1/n)$.

Proof: Notice that

$$\begin{aligned}(\sigma - s_n)^4 &= \left(\frac{\sigma - s_n}{\sqrt{\sigma^2 - s_n^2}}\right)^4 \left(\sqrt{\sigma^2 - s_n^2}\right)^4 \\ &= \left(\sqrt{\frac{\sigma - s_n}{\sigma + s_n}}\right)^4 \left(\sigma^2 - s_n^2\right)^2 \\ &\leq \left(\sigma^2 - s_n^2\right)^2\end{aligned}$$

which means that

$$\begin{aligned}\mathrm{E}(\sigma - s_n)^4 &\leq \mathrm{E}\left(\left(\sigma^2 - s_n^2\right)^2\right) \\ &= \mathrm{E}\left[\left(\mathrm{E}\left(X_1^2\right) - \mu^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2\right)\right]^2 \\ &= \mathrm{E}\left[\left(\mathrm{E}(X_1^2) - \frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \left(\mu^2 - \bar{X}_n^2\right)\right]^2 \\ &\leq 2\left[\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \mathrm{E}(X_1^2)\right)^2\right] + \mathrm{E}\left[\left(\bar{X}_n^2 - \mu^2\right)^2\right]\right] \\ &= 2\left[\frac{\mathrm{var}(X^2)}{n} + \frac{1}{n}\mathrm{E}\left[n(\bar{X}_n - \mu)^2(\bar{X}_n + \mu)^2\right]\right] \\ &\leq 2\left[\frac{\mathrm{E}(X^4)}{n} + \frac{1}{n}\sqrt{\mathrm{E}(\sqrt{n}(\bar{X}_n - \mu))^4 \mathrm{E}(\bar{X}_n + \mu)^4}\right]\end{aligned}$$

which is $O(1/n)$ because $\mathrm{E}(X)^4 < \infty$.□

Lemma 2: Let $X_1, X_2, \ldots X_n$ be iid random variables such that $\mathrm{E}(X_i)^4 < \infty$. If s_n denotes the usual sample standard deviation of $X_1, X_2, \ldots X_n$ and $\delta_n \to 0$, then $\Pr(s_n < \delta_n) = O(1/n)$.

Proof:

$$\begin{aligned}
\Pr(s_n < \delta_n) &= \Pr(s_n^2 < \delta_n^2) \\
&= \Pr\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2 < \delta_n^2\right) \\
&= \Pr\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \mathrm{E}(X_i^2)\right) - (\bar{X}_n^2 - \mu^2) < \delta_n^2 - \sigma^2\right) \\
&\leq \Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \mathrm{E}(X_i^2)\right| > \left|\frac{\sigma^2 - \delta_n^2}{2}\right|\right) \\
&\quad + \Pr\left(|\bar{X}_n^2 - \mu^2| > \left|\frac{\sigma^2 - \delta_n^2}{2}\right|\right) \\
&\leq \mathrm{E}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i^2 - \mathrm{E}(X_i^2)\right)^2 \left(\frac{4}{n\sigma^4}\right) + \mathrm{E}(|\bar{X}_n^2 - \mu^2|^2)\frac{4}{\sigma^4} \\
&= O(1/n)\square
\end{aligned}$$

Next we show that $\mathrm{var}(r_{nk}) \to 0$ and therefore each of the remaining covariance pieces converges to 0. Actually it suffices to show that $\mathrm{E}\left(r_{nk}^2\right) \to 0$. Using the Schwartz inequality

$$\begin{aligned}
\mathrm{E}\left(r_{nk}^2\right) &= \mathrm{E}\frac{n(\bar{X}_{nk} - \mu_k)^2}{\sigma_k^2}\left(\frac{\sigma_k}{s_{nk} \wedge \delta_n} - 1\right)^2 \\
&\leq \sqrt{\mathrm{E}\left(\frac{\sqrt{n}(\bar{X}_{nk} - \mu_k)}{\sigma_k}\right)^4}\sqrt{\mathrm{E}\left(\frac{\sigma_k}{s_{nk} \wedge \delta_n} - 1\right)^4} \qquad (3.8)
\end{aligned}$$

Now $\mathrm{E}\left(\frac{\sqrt{n}(\bar{X}_{nk}-\mu_k)}{\sigma_k}\right)^4$ is $O(1)$. To see this we rewrite it as:

$$\begin{aligned}
\mathrm{E}\left(\frac{\sqrt{n}(\bar{X}_{nk} - \mu_k)}{\sigma_k}\right)^4 &= \frac{n^2}{\sigma_k^4}\mathrm{E}(\bar{X}_{nk} - \mu_k)^4 \\
&= \frac{1}{n^2\sigma^4}\mathrm{E}\left(\sum_{i=1}^{n}(X_{ik} - \mu_k)^4\right).
\end{aligned}$$

But by Corollary 2, page 368 in Chow and Teicher (1988) this expected value is $O(n^2)$. So $\mathrm{E}\left(\frac{\sqrt{n}(\bar{X}_{nk}-\mu_k)}{\sigma_k}\right)^4$ is $O(1)$. Now we concentrate on the second piece of eq. (3.8).

$$\begin{aligned}
\mathrm{E}\left(\frac{\sigma_k-(s_{nk}\wedge\delta_n)}{s_{nk}\wedge\delta_n}\right)^4 &= \mathrm{E}\left(\left(\frac{\sigma_k-s_{nk}}{s_{nk}}\right)^4 : s_{nk}>\delta_n\right) \\
&\quad +\mathrm{E}\left(\left(\frac{\sigma_k-\delta_n}{\delta_n}\right)^4 : s_{nk}<\delta_n\right) \\
&\leq \frac{1}{\delta_n^4}\mathrm{E}\left(\sigma_k-s_{nk}\right)^4+\left(\frac{\sigma_k}{\delta}\right)^4\Pr\left(s_{nk}<\delta_n\right) \\
&= O\left(\frac{1}{n\delta_n^4}\right).
\end{aligned}$$

by Lemmas 1 and 2. Therefore this second term of eq. (3.8) goes to 0 which means that the variance and covariance terms in question go to 0 as $n\to\infty$. Therefore $\mathrm{cov}(\tilde{t}_k,\tilde{t}_{k'})\to\sigma_{kk'}/(\sigma_k\sigma_k')$.

Next we show that the covariance obtained from the bootstrap samples converges to the covariance of the original test statistics. By the strong law of large numbers $\tilde{\Sigma}^*_{Nn}$ converges to $\mathrm{cov}(\mathrm{t}^{*(1)}_{n\cdot})=\tilde{\Sigma}^*_n$, so we concentrate on the first bootstrap sample.

Let $\tilde{t}^*_{nk}$ be the kth test statistic and notice that we can write

$$\begin{aligned}
\tilde{t}^*_{nk} &= \frac{\sqrt{n}\left(\bar{Y}_{nk}-\bar{X}_{nk}\right)}{\sigma_k}\left(\frac{\sigma_k}{s^*_{nk}\wedge\delta_n}-1\right)+\frac{\sqrt{n}\left(\bar{Y}_{nk}-\bar{X}_{nk}\right)}{\sigma_k} \\
&= r^*_{nk}+p^*_{nk}
\end{aligned}$$

and similarly $\tilde{t}^*_{nk'}=r^*_{nk'}+p^*_{nk'}$. As before, we start with $\mathrm{E}\left(p^*_{nk}p^*_{nk'}|X\right)$.

$$\mathrm{E}\left(p^*_{nk}p^*_{nk'}|X\right) = \frac{n}{\sigma_k\sigma_{k'}}\mathrm{cov}\left(\bar{Y}_{nk},\bar{Y}_{nk'}|X\right)$$

$$\rightarrow \frac{\sigma_{kk'}}{\sigma_k \sigma_{k'}}$$

by Theorem 2.2 (a), Bickel and Freedman (1981). As before, we claim that the other terms converge to 0 and to prove this claim it suffices to show that $\mathrm{E}\left(r_{nk}^{*2}|X\right) \rightarrow 0$.

We start by applying the Schwartz inequality:

$$\begin{aligned}
\mathrm{E}\left(r_{nk}^{*2}|X\right) &= \mathrm{E}\left(\left(\frac{\sqrt{n}\left(\bar{Y}_k - \bar{X}_k\right)}{\sigma_k}\right)^2 \left(\frac{\sigma_k}{s_{nk}^* \wedge \delta_n} - 1\right)^2 |X\right) \\
&\leq \sqrt{\mathrm{E}\left(\left(\frac{\sqrt{n}\left(\bar{Y}_k - \bar{X}_k\right)}{\sigma_k}\right)^4 |X\right)} \sqrt{\mathrm{E}\left(\left(\frac{\sigma_k}{s_{nk}^* \wedge \delta_n} - 1\right)^4 |X\right)}.
\end{aligned}$$

By Corollary 2, page 368 of Chow and Teicher (1988) the first term is $O(1)$ so we concentrate on the second term.

$$\begin{aligned}
\mathrm{E}\left(\left(\frac{\sigma_k}{s_{nk}^* \wedge \delta_n} - 1\right)^4 |X\right) &= \mathrm{E}\left(\left(\frac{\sigma_k - s_{nk}^*}{s_{nk}^*}\right)^4 : s_{nk}^* > \delta_n, X\right) \\
&\quad + \mathrm{E}\left(\left(\frac{\sigma_k - \delta_n}{\delta_n}\right)^4 : s_{nk}^* < \delta_n, X\right) \\
&\leq \frac{1}{\delta_n^4} \mathrm{E}\left(\left(\sigma_k - s_{nk}^*\right)^4 |X\right) \\
&\quad + \left(\frac{\sigma_k}{\delta_n}\right)^4 \Pr\left(s_{nk}^* < \delta_n |X\right).
\end{aligned}$$

Now, since $\mathrm{E}(Y_{ik}^4) \rightarrow \infty$ we can use Lemma 2 to see that $\Pr\left(s_{nk}^* < \delta_n |X\right) = O(1/n)$. We can rewrite $\mathrm{E}\left(\left(\sigma_k - s_{nk}^*\right)^4 |X\right)$ as:

$$\mathrm{E}\left(\left(\sigma_k - s_{nk}^*\right)^4 |X\right) = \mathrm{E}\left(\left(\left(\sigma_k - s_{nk}\right) + \left(s_{nk} - s_{nk}^*\right)\right)^4 |X\right).$$

Using Minkowski's inequality we get

$$\left[\mathrm{E}\left(\left(\sigma_k - s_{nk}^*\right)^4 |X\right)\right]^{1/4} \leq \left[\mathrm{E}\left(\left(\sigma_k - s_{nk}\right)^4 |X\right)\right]^{1/4}$$

$$+\left[\mathrm{E}\left(\left(s_{nk}-s_{nk}^{*}\right)^{4}|X\right)\right]^{1/4}$$
$$= O\left(\frac{1}{n^{1/4}}\right)+O\left(\frac{1}{n^{1/4}}\right)$$

by Lemma 1. Therefore $\mathrm{E}\left(\left(\sigma_k - s_{nk}^{*}\right)^4 |X\right) = O(1/n)$ and

$$\mathrm{E}\left(\left(\frac{\sigma_k}{s_{nk}^{*}\wedge\delta_n}-1\right)^4 |X\right)\to 0$$

because $n\delta_n^4 \to \infty$. Therefore $\mathrm{cov}(\tilde{t}_k^{*}, \tilde{t}_{k'}^{*}) \to \frac{\sigma_{kk'}}{\sigma_k\sigma_{k'}}$ which is the covariance of the original test statistics.

We have shown that

1. $\tilde{\Sigma}_{Nn}^{*} - \tilde{\Sigma}_{n}^{*} \to 0$ as $N \to \infty$.

2. $\tilde{\Sigma}_{n}^{*} - \Sigma \to 0$ in probability as $n \to \infty$.

But, notice also that since $\Pr(s_{nk}^{*} < \delta_n) \to 0$

$$\Sigma_{Nn}^{*} - \tilde{\Sigma}_{Nn}^{*} \to 0$$

as $n \to \infty$ where Σ_{Nn}^{*} is the estimated covariance matrix based on the usual bootstrap t-statistics. Therefore, the final conclusion is that the bootstrapped estimate of the covariance matrix computed from the usual t-statistics, t_{nk}^{*}, is a consistent estimator for the true covariance matrix of the original test statistics.

3.12 Appendix B. Asymptotic Distribution of S_W

In this appendix we prove that the statistic S_W has an asymptotic distribution which is a noncentral χ_r^2 distribution with r being the rank of the covariance matrix for the individual statistics.

Let $\underline{y}$ be distributed multivariate normal with mean vector $\underline{\mu}$ and covariance matrix $\mathbf{\Sigma}$, where $\mathbf{\Sigma}$ is $p \times p$. Assume that the rank of $\mathbf{\Sigma}$ is known to be $r < p$. Consider the spectral decomposition,

$$\mathbf{\Sigma} = \mathbf{Q}\Lambda\mathbf{Q}'$$

where Λ is a diagonal matrix containing the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_p$ of $\mathbf{\Sigma}$, and $\mathbf{Q}$ is the matrix containing the normalized eigenvectors $q_1, q_2, ..., q_p$ corresponding to $\lambda_1, \lambda_2, ..., \lambda_p$. Since the rank of $\mathbf{\Sigma}$ is $r < p$, when the eigenvalues are ordered as $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$, the first r of the eigenvalues are positive and the remaining $p-r$ eigenvalues are zero. This allows us to partition the matrices Λ and $\mathbf{Q}$ as

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix}$$

where $\mathbf{Q}_1$ is a $p \times r$ matrix and Λ_1 is $r \times r$.

We begin the process of finding the asymptotic distribution of S_W by looking at the distribution of $\underline{y}'(\mathbf{Q}_1\Lambda_1^{-1}\mathbf{Q}_1')\underline{y}$. To make notation a bit easier, let $\mathbf{A} = \mathbf{Q}_1\Lambda_1^{-1}\mathbf{Q}_1'$. By Corollary 2s.1 in Searle (1971, page 69) $\underline{y}'\mathbf{A}\underline{y}$ has a χ^2_r distribution with noncentrality parameter $\frac{1}{2}\underline{\mu}'\mathbf{A}^{-1}\underline{\mu}$.

The problem of the asymptotic distribution of S_W is a bit more complicated, however, because the true covariance matrix $\mathbf{\Sigma}$ is usually unknown and is estimated by the matrix $\mathbf{S}$. The following argument is similar to that used in the proof of Theorem 2 by Amemiya and Fuller (1984). As n (the number of observations) goes to infinity, $\mathbf{S}$ converges in probability to $\mathbf{\Sigma}$. Therefore, there exists a subsequence $\{n_i\}$ over which $\mathbf{S}$ converges almost surely to $\mathbf{\Sigma}$. Let ω be a point in the probability

space of all sequences of observations. Fix ω such that $\mathbf{S}$ converges to $\mathbf{\Sigma}$. Franklin (1968, pg. 191) shows that the $\hat{\lambda}_i$ are locally continuous functions of the elements of $\mathbf{S}$ which implies that $\hat{\lambda}_i(\omega)$ converges to $\lambda_i(\omega)$.

For each eigenvalue, $\hat{\lambda}_i$, there is a corresponding normalized eigenvector, $\underline{\hat{q}}_i$. Because $\underline{\hat{q}}_i$ is normalized, the elements in $\underline{\hat{q}}_i$ are bounded in absolute value by 1 and there exists a subsequence over which $\underline{\hat{q}}_i$ converges to some $\underline{q}_i(\omega)$ almost surely. This means that $\mathbf{S}_n\underline{\hat{q}}_i$ converges to $\mathbf{\Sigma}\underline{q}_i(\omega)$ and $\hat{\lambda}_i\underline{\hat{q}}_i$ converges to $\lambda_i\underline{q}_i(\omega)$ for the fixed ω. Finally, since

$$\mathbf{S}_n\underline{\hat{q}}_i = \hat{\lambda}_i\underline{\hat{q}}_i,$$

we have that

$$\mathbf{\Sigma}\underline{q}_i(\omega) = \lambda_i\underline{q}_i(\omega)$$

implying that $\underline{q}_i(\omega)$, the limit of the eigenvectors of the $\mathbf{S}$ are indeed the eigenvectors of $\mathbf{\Sigma}$. This means that $\hat{\mathbf{Q}}_1 = (\hat{q}_1, \hat{q}_2, ..., \hat{q}_r)$ converges to $(q_1(\omega), q_2(\omega), ..., q_r(\omega)) = \mathbf{Q}_1(\omega)$, and since the eigenvalues converge almost surely to the true eigenvalues, $\hat{\Lambda}_1^{-1}$ converges to Λ_1^{-1}. Now, the limit of each eigenvector does depend on the ω chosen but will always be of the form $\underline{q}_i\mathbf{R}_\omega$ where $\mathbf{R}_\omega$ is an orthogonal matrix and $\underline{q}_i$ is the normalized eigenvector of $\mathbf{\Sigma}$ corresponding to λ_i. This means that for any ω

$$\begin{aligned}
\hat{\mathbf{Q}}_1\hat{\Lambda}_1^{-1}\hat{\mathbf{Q}}_1' &\rightarrow \mathbf{Q}_1(\omega)\Lambda_1^{-1}\mathbf{Q}_1(\omega)' \\
&= \mathbf{Q}_1\mathbf{R}_\omega\Lambda_1^{-1}\mathbf{R}_\omega'\mathbf{Q}_1' \\
&= \sum \underline{q}_i\mathbf{R}_\omega\frac{1}{\lambda_i}\mathbf{R}_\omega'\underline{q}_i' \\
&= \sum \frac{1}{\lambda_i}\underline{q}_i\underline{q}_i'
\end{aligned}$$

because $\mathbf{R}_\omega$ is orthogonal. This means that the limit is no longer dependent on ω, i.e. the limit of $\hat{\mathbf{Q}}_1\hat{\Lambda}_1^{-1}\hat{\mathbf{Q}}_1'$ is the same for all ω such that $\mathbf{S}(\omega)$ converges to $\mathbf{\Sigma}(\omega)$. But

the probability of this set is 1 so $\hat{\mathbf{Q}}_1\hat{\Lambda}_1^{-1}\hat{\mathbf{Q}}_1'$ converges to $\mathbf{Q}_1\Lambda_1^{-1}\mathbf{Q}_1'$ with probability 1.

Since $\underline{t}$ converges in distribution to a multivariate normal with mean $\underline{\mu}$ and covariance matrix $\mathbf{\Sigma}$, and $\hat{\mathbf{Q}}_1\hat{\Lambda}_1^{-1}\hat{\mathbf{Q}}_1'$ converges to $\mathbf{Q}_1\Lambda_1^{-1}\mathbf{Q}_1'$ with probability 1, we have the desired result that

$$\underline{t}'\hat{A}\underline{t} \rightarrow \chi^2_r\left(\frac{1}{2}\underline{\mu}'A\underline{\mu}\right)$$

in distribution.

4. CONCLUSION

The goal of this work has been to improve the method of analysis currently in use (e.g., Kernan and Meeker, 1992) for rat behavior data which can be characterized as categorical temporal data. Experiments using control and treated rats result in vectors of behaviors of the rats as they occur during a 900 second period. Researchers have traditionally measured many different aspects of this data and tried to make the single conclusion of whether a difference, not attributable to normal variability, exists between the behaviors of the two groups of rats. The method used as a starting point for this research appears in Kernan and Meeker (1992) and involves measuring the many aspects of interest from the data, performing a hypothesis test for each aspect and counting the number of significant tests. Two areas of potential improvement to this method were identified and each area is discussed in detail and improved upon in this dissertation.

The first area identified for improvement is the group of measurements made on the data using a temporal analogue to the spatial statistic, the K function. This statistic was first introduced to this type of experiment by Kernan et al. (1988). The K function is a relatively new statistic which is used to detect temporal patterns in data and researchers continue to study its estimators. Since the publication of Kernan et al. (1988), spatial statisticians have improved upon the estimator of the K

function used in Kernan et al. (1988); the most recent improvement for spatial data appears in Stein (1993). We derived the temporal analogue to Stein's estimator for use with the rat behavior data.

We also improved the individual hypothesis test involving the K function, making it a more sensitive test. Kernan and Meeker (1992) chose eight points along the continuous K function to test the hypothesis that the functions for the two groups were equal. We created a set of two tests that use the whole function within a specific region to compare, within pairs of rats, the differences between the rat's estimate and a reference function provided by the null hypothesis. By incorporating the whole function into the test, we use more information and hence have a more sensitive test.

While we have improved the application of the K function to this type of data, more work remains to be done in this area. Kernan et al. (1988) also used a statistic based on the cross-K function to test for temporal pattern differences for pairs of acts. A derivation similar to that for the temporal analogue to the K function should be possible and the estimator of the cross-K function updated. Also, tests using the cross-K function should be developed, most likely along the same lines as the tests involving the K function presented here.

The second area identified for improvement in the analysis involves the final decision making process of whether the two groups of behaviors are the same or not. Kernan and Meeker (1992) used a statistic which was a count of significant tests, denoted here by S_B. We suggested two other options: a sum of the squared test statistics, denoted by S_T; and a Wald-like combination of the test statistics, denoted by S_W.

In general we expected S_T to perform, at least slightly, better than S_B because

S_T accounts for the strength of the significance in individual tests whereas S_B merely notes the fact of significance. This expectation was indeed found to be true under the models using the covariance matrices estimated from each of three data sets used. S_T was generally more powerful than S_B, but only slightly.

The other expectation was that S_W would do much better than either S_B or S_T because the test statistics defined by this experiment are obviously correlated to some extent and neither S_B nor S_T account for this correlation. Indeed, for the models used for two of the three data sets S_W was found to be much more powerful then either S_B or S_T. For the third model tested, S_W performed about the same as S_B and S_T. The basic difference among the three models is the covariance matrix governing the test statistics. For the two models with relatively large non-centrality parameters, S_W was more powerful than S_B and S_T. When the noncentrality parameter was closer to 0, S_W performed similarly to S_B and S_T.

Based on the three data sets used for this analysis, it appears that when the rats are more active, either because of the treatment received or because they are observed under red light, the noncentrality parameter based on the estimated covariance matrix is larger and therefore S_W has more power. Typically the researcher will not know how much activity to expect from the rats in an experiment, and therefore will not be able to predict a covariance matrix in advance, so we suggest using S_W for all cases. It does not appear to be much worse than S_B or S_T and is possibly much more powerful for detecting differences.

The improvements made in this work to the current method of analysis are by no means the end. Several questions have arisen throughout this project which merit consideration. The first question is how the length of the observation period

would change the power of the statistics. Most likely a shorter period of observation time would not be preferred because more variability would be introduced into the data. And, 15 minutes was chosen because, in practice, once rats become used to an environment they reduce their level of activity so much that little more information could be gained. But, perhaps after 15 minutes they could be moved to another novel environment and observed for a longer time period.

One problem that arose in this research is that there was much rat-to-rat variability. In order to reduce this variability somewhat, we suggest, when possible, observing both the controls and the treated rats before administering the treatment and then observing them all again after the treatment. This is not always possible because frequently the treatment is given to pregnant rats and the offspring are observed. Also, the question of the novel environment occurs. If the rats are put into this environment as controls and then again as treated rats, the second time the environment is no longer novel. For this type of design, a second novel environment must be constructed that is similar enough not to give an environment effect, but different enough to be novel to the rats.

A third question is which test statistics to include in the overall statistic. In this article we have addressed combining information from several applications each of three types of statistics. Kernan and Meeker (1992) used four types in their analysis. We have not done a comprehensive study on which individual statistics to include. There may be unnecessary redundancy or there may be other statistics, not yet considered, that should be included.

BIBLIOGRAPHY

Baumeister, A.A. (1978) Origins and Control of Stereotyped Movements. Quality of Life in Severely and Profoundly Mentally Retarded People: Research Foundations for Improvement. Washington D.C., American Association in Mental Deficiency, Monograph #3, 353-384.

Diggle, P.J. (1983) *Statistical Analysis of Spatial Point Patterns.* New York: Academic Press.

Kernan, W. J., and Meeker, W. Q. (1992) "A Statistical Test to Assess Changes in Spontaneous Behavior of Rats Observed with a Computer Pattern Recognition System." *Journal of Biopharmaceutical Statistics*, 2, 115–135.

Kernan, W. J. Jr., Mullenix, P. J., and Hopper, D. L. (1989) "Time Structure Analysis of Behavioral Acts Using a Computer Pattern Recognition System." *Pharmacology Biochemistry and Behavior*, 34, 863–869.

Kernan, W. J. Jr., Mullenix, P. J., Kent, R., Hopper, D. L., and Cressie, N. A. C. (1988) "Analysis of the Time Distribution and Time Sequence of Behavioral Acts." *International Journal of Neuroscience*, 43, 35–51.

Mullenix, P. J., Kernan, W. J., Tassinari, M. S., and Schunior, A. (1989) "Generation of dose-response data using activity measures." *Journal of the American College of Toxicology*, 8, 185-197.

Norton, S., Mullenix, P., and Culver, B. (1976) "Comparison of the Structure of Hyperactive Behavior in Rats after Brain Damage from X-irradiation, Carbon Monoxide and Pallidal Lesions." *Brain Research*, 116, 49–67.

Pohl, P. (1976) "Spontaneous Fluctuation in Rate of Body Rocking: A Methodological Note". *Journal of Mental Deficiency Research*, 20, 61-65.

Pontius, A. A. (1973) "Dysfunction Patterns Analogous to Frontal Lobe System and Caudate Nucleus Syndrome in some Groups of Minimal Brain Dysfunction." *Journal of American Medical Women's Association*, 28, 285-291.

Stein, M.L. (1993) "Asymptotically Optimal Estimation for the Reduced Second Moment Measure of Point Processes." *Biometrika*, 80, 443-449.

Wender, P. H. (1971) *Minimal Brain Dysfunction in Children.* New York: Wiley - Interscience.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation for both major professors: Dr. William Meeker and Dr. Noel Cressie. Both have been extremely supportive of me throughout the process of this project. Any mistakes are mine, not theirs. I would also like to thank the members of my committee for their help: Dr. William Kernan, Dr. Kenneth Koehler and Dr. Robert Stephenson. Dr. Kernan deserves a special mention for supplying me with the problem at hand, several data sets, and quite a bit of background information about the experiment. Also I would like to thank two other members of the Iowa State Statistics Department, Dr. Yasuo Amemiya and Dr. Krishna Athreya, for listening to my thoughts and suggesting improvements.

A very large thank you goes to Kathy Shelley and Bud Meador, Iowa State Department of Statistics, and John Hascall of the Iowa State University Computation Center. All three helped to keep my computations running. This project required a total of about 6000 hours of cpu time and without Kathy, Bud and John I might still be trying to compute the results.

Finally I want to thank my husband, Scott, and the rest of my family. Each has listened to me and tried in their own way to relieve my stress as much as possible. I could not have accomplished this task without their support and understanding.