A JOINT INTEGRAL TRANSFORMATION APPROACH TO GOODNESS OF FIT

By

CHARLES CLEVELAND STEWART

Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1963

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1966

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
July, 1971

A JOINT INTEGRAL TRANSFORMATION
APPROACH TO GOODNESS OF FIT

Thesis Approved:

J. Leroy Folks
Thesis Adviser

David L. Weeks

Jeanne Agnew

James E. Shamblin

D. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to Dr. J. Leroy Folks for suggesting the thesis topic and for serving as chairman of my advisory committee. His comments, guidance, and especially, long patience, were deeply appreciated. I also wish to express my appreciation to Dr. David L. Weeks, Dr. Jeanne L. Agnew, and Dr. James E. Shamblin for serving as members of my committee.

Acknowledgment is particularly due to those responsible for the National Defense Education Act Fellowships. Special thanks also goes to Mrs. Mary Bonner for the diligent work in the typing and preparation of this thesis.

Last, but not least, I wish to express the deepest appreciation to my wife, Dixie, for all those things which cannot be said with words.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Goodness of fit tests are concerned with testing that the probability law or distribution function of a sampled population is of a specific form, for example, normal with mean ten and standard deviation two. The broader problems of testing that a sampled population has a distribution function belonging to a parametric family, and that two or more sampled populations have the same distribution function are also classified as goodness of fit problems.

This study is primarily concerned with constructing and evaluating test statistics for the simple goodness of fit hypothesis; that is, the population distribution function $F(\cdot)$ has a completely specified form $F_0(\cdot)$. The composite goodness of fit hypothesis, that is, specifying the form of the distribution function only up to certain unknown parameters, is discussed briefly in Chapter II and then again in Chapter VI. The problem of testing that two or more population distribution functions are the same is not considered in this study.

The first analytic procedure for testing goodness of fit was given by Karl Pearson (35) in 1900, which is relatively early in the history of statistics. Thus this problem is a classic in the field and the literature is vast. Even the literature on the Pearson procedure is very extensive. A brief review of the literature is given in Chapter II.

Notation and Definitions of Terms

Before stating the goodness of fit problem a brief review of theory of significance testing is necessary. Let X be a random variable with sample space R_n (n-dimensional Euclidean space), and probability distribution function $F_\theta(\cdot)$ where θ is a parameter belonging to some parameter space Ω. The null hypothesis is then

$$H_0 : \theta \in \Omega_0 , \text{ where } \Omega_0 \subset \Omega ,$$

versus the alternative

$$H_A : \theta \in \Omega_A , \text{ where } \Omega_A \subset \Omega$$

and

$$\Omega_0 \cap \Omega_A = \emptyset .$$

Any statistic $T^*(X)$ with values in R_1 can be considered a test statistic for this hypothesis problem if its probability distribution function $G_\theta(\cdot)$ is completely specified for $\theta \in \Omega_0$. The obvious purpose of constructing a test statistic is to measure departure from the null hypothesis in the direction of the alternative. For practical problems it is usually possible to construct test statistics such that, say, small values clearly measure departure from the null hypothesis in the direction of the alternative (30).

Accordingly, assume that $T^*(X)$ is chosen so that small values are more consistent with the alternative hypothesis. If x is an observed value of X, $t^*(x)$ an observed value of $T^*(X)$, then the statistical or significance test consists of computing the observed significance level given by

$$\ell(t^*) = P_0[T^*(X) \le t^*(x)] = G_0(t^*) ,$$

where $G_0(\cdot)$ is the (completely specified) distribution function of $T^*(X)$ for $\theta \in \Omega_0$. Clearly small values of the significance level are more consistent with the alternative hypothesis.

The observed significance level obviously satisfies

$$0 \le \ell(t^*) \le 1 .$$

However, not all intermediate values are necessarily achievable. For example, if $T^*(X)$ is a random variable of the discrete type then the significance level of T^* can achieve at most countably many values. If T^* is of the continuous type for $\theta \in \Omega_0$, then $G_0(\cdot)$ is continuous and $\ell(t^*)$ achieves all values in $(0, 1)$.

The observed significance level $\ell(t^*)$ is itself the realization of the random variable

$$L(T^*) = G_0(T^*)$$

with distribution function, for $\theta \in \Omega$ and achievable $\ell \in (0, 1)$,

$$\begin{aligned} H^*(\ell) &= P_\theta[L(T^*) \le \ell] \\ &= P_\theta[G_0(T^*) \le \ell] \\ &= P_\theta[T^* \le t^*_\ell] \end{aligned}$$

where

$$t^*_\ell = \inf_{-\infty < t < \infty} \{t^* \mid G_0(t^*) = \ell\} .$$

Only achievable significance levels will be considered. If $\theta \in \Omega_0$ then $H^*_\theta(\ell)$ will be written $H^*_0(\ell)$, and $H_0(\ell) = \ell$ for each achievable ℓ.

If T^* is of the continuous type for $\theta \in \Omega_0$, then

$$H_0^*(\ell) = \ell, \quad \text{for all} \quad \ell \in (0,1) ;$$

that is $L(T^*)$ has the uniform distribution on $(0, 1)$.

If $\theta \in \Omega_A$ the value of $H_\theta^*(\ell)$ is called the sensitivity or power of T^* at significance level ℓ and parameter value θ. Though many criteria have been devised for comparing test statistics, it is generally agreed that one should attempt to choose a test with good sensitivity. Several definitions are necessary for definiteness.

Definition 1.1: Two test statistics T_1 and T_2 are said to be comparable if they have the same set of achievable significance levels.

Definition 1.2: If the alternative hypothesis is simple (Ω_A contains one element θ_1), and if T_1 and T_2 are comparable statistics, then T_1 is said to be more sensitive than T_2 if

$$H_{\theta_1}^{(1)}(\ell) \geq H_{\theta_1}^{(2)}(\ell)$$

for all achievable $\ell \in [0, 1]$, with strict inequality holding for at least one $\ell \in (0, 1)$.

Definition 1.3: If a statistic T is a most sensitive test for all $\theta \in \Omega_A$, then T is said to be a uniformly most sensitive test statistic.

Definition 1.4: A test statistic T^* is said to be unbiased if

$$H_\theta^*(\ell) \geq \ell$$

for all achievable $\ell \in [0, 1]$ and all $\theta \in \Omega_A$.

In most testing problems the random variable X is of the form

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix}$$

where the X_i, $i = 1, \ldots, n$, have identical distribution functions and are independent; that is, the observed value x is a random sample of sample size n . It is of interest to inquire whether a test defined for each sample size n has good properties when n is large. One such property, consistency, is defined below. Another, Bahadur (2) exact slope, is discussed in Chapter IV.

Definition 1.5: A sequence of test statistics $\{T^{(n)}\}$ is consistent for the alternative if, for each $\theta \epsilon \Omega_A$,

$$\lim_{n\to\infty} H_\theta^{(n)}(\ell) = 1$$

where the convergence is point-wise for each $\ell \epsilon (0, 1)$.

No attempt will be made here to describe procedures for constructing unbiased or most sensitive test statistics. The reader is referred to Lehmann (26), Fraser (17), Finley (13), or Moore (32) for such procedures.

Simple Goodness of Fit Hypothesis

The goodness of fit problem of primary interest for this study can be stated as follows: Let S be the set of all admissible distribution functions of a random variable X with values in R_1, and let S_A be a subset of S not containing $F_0(\cdot)$. Given a random sample $x_1, x_2, \ldots, x_n$, realizations of the independent random variables $X_1, X_2, \ldots, X_n$ each with the same unknown distribution function $F(\cdot) \in S$, test the null hypothesis

$$H_0: F(x) \equiv F_0(x), \quad x \in R_1,$$

versus the alternative

$$H_A: F(x) \in S_A, \ x \in R_1.$$

Because it is difficult, from a practical point of view, to select the admissible set S to be anything less than, say, the set of all continuous distribution functions and the alternative set $S_A = S - \{F_0\}$, it is not possible to use the usual parametric methods for constructing "good" test statistics. (Lehmann (26)). Test statistics are usually constructed to be intuitively satisfying in the sense that they will provide some sensitivity to a wide class of alternatives. There are usually great mathematical difficulties connected with finding the exact sensitivity of goodness of fit tests. There are, however, procedures available for obtaining good approximations. One such procedure, synthetic sampling, is used in Chapter V.

There are, for some of the most well-known procedures, some mathematical difficulties encountered in finding the exact distribution function assuming the null hypothesis. (See Chapter II). For such test statistics approximations are necessary for computing the significance level. A test T^* is said to be exact if the exact form of the distribution function of T^* is known assuming the null hypothesis is true.

When the X_i, $i = 1, 2, \ldots, n$, are of the continuous type each with distribution function F_0, then it is well-known that U_i, $i = 1, 2, \ldots, n$, defined by

$$U_i = F_0(X_i), \quad i = 1, 2, \ldots, n,$$

are independent, each with the uniform distribution on the interval zero to one. Suppose S is the set of all continuous univariate distribution functions (distribution functions of all random variables of the continuous type) and $S_A = S - \{F_0\}$. Suppose we always perform this "integral transformation." No matter what the form of F_0, the application of the transformation reduces the hypothesis problem H_0 versus H_A to the problem

$$H_0' : G(u) = G_0(u), \quad -\infty < u < \infty$$

versus

$$H_A' : G(u) \neq G_0(u), \quad -\infty < u < \infty$$

where

$$G_0(u) = \begin{cases} 0, & u < 0 \\ u, & 0 \leq u \leq 1 \\ 1, & u > 1 . \end{cases}$$

This transformation gives a "nonparametric" character to any procedure based on the random variables $U_1, U_2, \ldots, U_n$.

Composite Goodness of Fit Hypothesis

In the composite case, the null hypothesis specifies only that $F(\cdot)$ is a member of a certain parametric class $C = \{F_0(\cdot;\theta), \theta \in \Omega\}$.

Typically C is the class of normal distribution functions and $\theta = (\mu, \sigma^2)$, μ the mean and σ^2 the variance. Two methods for reducing the composite hypothesis to a simple hypothesis are given in Chapter II. A method for reducing the composite hypothesis to H_0' and H_A' is given in Chapter VI.

Combining Independent Significance Levels

Again let $X_1, X_2, \ldots, X_p$ be independent continuous random variables, but now assume that the distribution functions are $F_1(\cdot), F_2(\cdot), \ldots, F_p(\cdot)$, respectively. If a modification to the integral transformation is applied, so that

$$U_i = F_i(X_i) , \quad i = 1, 2, \ldots, p ,$$

the p random variables $U_1, U_2, \ldots, U_p$ will again be independent uniform random variables.

Now consider the problem of combining independent tests of significance. Let T_i, $i = 1, 2, \ldots, p$, be independent test statistics for testing the null hypothesis

$$H_{0,i}: \theta_i \in \Omega_{0,i} , \quad i = 1, 2, \ldots, p ,$$

versus the alternatives

$$H_{A,i}: \theta_i \in \Omega_{A,i} , \quad i = 1, 2, \ldots, p .$$

That is, T_1 is a test statistic for testing $H_{0,1}$ versus $H_{A,1}$, T_2 for testing $H_{0,2}$ versus $H_{A,2}$, etc. It is desired to construct a function of $T_1, T_2, \ldots, T_p$ that may be used to test the combined null hypothesis

$$H_0: \theta_i \in \Omega_{0,i}\ , \quad i = 1, 2, \ldots, p\ ,$$

versus the alternative

$$H_A: \text{at least one } \ \theta_i \in \Omega_{A,i}\ , \quad i = 1, 2, \ldots, p\ .$$

Suppose that each T_i is of the continuous type, and small values are taken to be consistent with the alternative $H_{A,i}$. Let $F_{0,i}(\cdot)$ represent the distribution function of T_i when $\theta_i \in \Omega_{0,i}$. Then the significance levels given by

$$L_i = F_{0,i}(T_i)\ , \quad i = 1, 2, \ldots, p\ ,$$

will be mutually independent uniform variables assuming H_0 true. If each of the T_i is an unbiased test statistic (Definition 1.4) of $H_{0,i}$ versus $H_{A,i}$, then $H_{\theta_i}^{(i)}(\ell) \geq \ell$ for all $\ell \in [0, 1]$ and $\theta_i \in \Omega_A$, where $H_{\theta_i}^{(i)}(\cdot)$ represents the distribution function of L_i. Even if some of the T_i are not unbiased, small values of the levels are taken to be consistent with the alternative. In either case it is reasonable to state a reduced hypothesis in terms of the significance levels as follows:

Given the random sample $\ell_1, \ell_2, \ldots, \ell_p$ (observed significance levels) test the null hypothesis

$$H_0'': H_0^{(i)}(\ell) = \begin{cases} 0\ , & \ell < 0 \\ \ell\ , & 0 \leq \ell \leq 1\ , \quad i = 1, 2, \ldots, p\ , \\ 1\ , & \ell > 1 \end{cases}$$

versus the alternative

$$H_A'': H_{\theta_i}^{(i)}(\ell) \begin{cases} = 0\ , & \ell < 0 \\ \geq \ell\ , & 0 \leq \ell \leq 1\ , \quad i = 1, 2, \ldots, p\ . \\ = 1\ , & \ell > 1 \end{cases}$$

Thus the problem of combining independent tests can be reduced to a "one-sided," simple goodness of fit problem.

Now the observed values of the test statistics, $t_1, t_2, \ldots, t_p$ arise from data $x_1, x_2, \ldots, x_p$, respectively. Presumably, one would base a test of H_0 either on the combined data $x = (x_1, x_2, \ldots, x_p)$ or on $t = (t_1, t_2, \ldots, t_p)$. It is assumed here, however, that either

(i) the values or else the forms of the distributions of x and t are unknown,

(ii) or this information is available but the distributions are such that there is no known or reasonably convenient method for constructing a single test of H_0 based on x or t.

Numerous examples of such problems have been given in the literature. (For example see Graybill (19) and Rao (38)). The problem of combining independent significance levels is considered first (Chapters III and IV). The general goodness of fit problem is considered in Chapters V and VI.

CHAPTER II

BRIEF REVIEW OF LITERATURE

Goodness of fit problems have been the subject of almost continuous research since Pearson's (33) test appeared in 1900. Therefore, a complete review of the literature would be a study in itself. It is necessary to include a brief review as a source of reference for comparative studies made in subsequent chapters. A more complete review of goodness of fit tests has been given by David (11). A description of procedures for combining independent test statistics has been given in a recent paper by van Zwet and Oosterhoff (47).

Simple Goodness of Fit Hypothesis

Pearson χ^2 Test

The test proposed by Pearson is commonly called the chi-squared test. To apply the test, one first divides the range of X into k disjoint intervals $I_j = (a_j, b_j]$, $j = 1, 2, \ldots, k$. Then the proportion of the null population, p_j, $j = 1, 2, \ldots, k$, associated with each interval is computed; i.e.,

$$p_j = F_0(b_j) - F_0(a_j) = P_0[a_j < X \leq b_j], \; j = 1, 2, \ldots, k .$$

The expected number of observations in each interval, E_j, is then given by $E_j = n\,p_j$ where n is the sample size. After the random sample has been collected, the observed number of observations in each interval, O_j , is tabulated. The chi-squared statistic is given by

$$CS = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j} .$$

The statistic derives its name from the fact that the limiting null distribution of CS is a chi-squared with $k - 1$ degrees of freedom. This fact was first demonstrated by Pearson. The approximate significance level of CS is given by

$$\ell_{CS} = P[\chi^2(k-1) \geq cs]$$

where $\chi^2(k-1)$ denotes a chi-squared variable with $k - 1$ degrees of freedom and cs the observed value of CS.

The larger the E_j , the better the approximation. Rules on the size of E_j are given in most textbooks. For example Cramer (9) suggests that if $E_j \geq 10$, $j = 1, 2, \ldots, k$, the approximation is sufficient for ordinary purposes. However, in a recent synthetic sampling study by Kempthorne (20) it is shown that the approximation is not seriously affected, for practical purposes, if the E_j are all 1 and n is as small as 10.

The χ^2 test is the most versatile of all procedures. The random variable X can be either of the continuous type or discrete type. It can also be applied to the composite goodness of fit hypothesis (14) and to multivariate problems (20). The χ^2 test is unbiased and consistent (24) and the sensitivity can be computed by means of the noncentral chi-squared distribution (8).

When the intervals I_j , $j = 1, \ldots, k$, are chosen so that $p_j = 1/k$, $j = 1, \ldots, n$, the χ^2 statistic becomes

$$CS^* = \frac{k}{n} \sum_{j=1}^{n} O_j^2 - n$$

where O_j is the observed number of observations in the j^{th} interval. This form is used in Chapter V for comparison with test statistics developed in this study.

Kolmogorov-Smirnov Test

Another well-known test is the Kolmogorov-Smirnov test which was first suggested by Kolmogorov (23). It bears Smirnov's name because Smirnov (44) gave an alternative derivation of the limiting null distribution and tabulated this function. To describe this test another definition is required.

Definition 2.1: Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables and let the order statistics be denoted by

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)} .$$

The empirical distribution function is defined as

$$F_n(x) = \begin{cases} 0 , & \text{for } x < X_{(1)} \\ i/n , & \text{for } X_{(i)} \leq x < X_{(i+1)} , \quad i = 1, 2, \ldots, n-1 \\ 1 , & \text{for } X_{(n)} \leq x . \end{cases}$$

The Kolmogorov-Smirnov statistic is

$$KS = \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)| ;$$

that is, the greatest absolute vertical discrepancy between the random function $F_n(\cdot)$ and the hypothesized distribution function $F_0(\cdot)$. When X is of the continuous type the exact null distribution of Ks has been tabulated (see Birnbaum (5) for numerical tabulation). The level of KS is given by

$$\ell_{KS} = P_0[KS \geq ks]$$

where ks is the observed value of KS.

An asymptotic expression for the greatest lower bound on the sensitivity has been given by Massey (31). Massey also pointed out that KS is a biased test for some alternatives. Fisz (16) proved that KS is consistent for a continuous alternative $G(\cdot)$ that satisfies the relation

$$\sup_{-\infty<x<\infty} |F_0(x) - G(x)| = \delta > 0 ,$$

When the probability integral transformation $U_i = F_0(X_i)$, $i = 1, 2, \ldots, n$, is applied to the random sample X_i, $i = 1, 2, \ldots, n$, then the test statistic can be expressed as

$$KS = \max\left[\max_{i=1,2,\ldots,n} \left(\frac{i}{n} - U_{(i)}\right), \max_{i=1,2,\ldots,n} \left(U_{(i)} - \frac{i-1}{n}\right)\right]$$

where $0 \leq U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n)} \leq 1$ denote the ordered transformed variables. A modification of the Kolmogorov-Smirnov statistic given by

$$KS^* = \max_{i=1,2,\ldots,n} \left|U_{(i)} - \frac{i}{n}\right|$$

was used by Shapiro, Wilk, and Chen (42) for comparisons of sensitivity among several statistics. Some results of this study are given in Chapter V.

Cramér-Von Mises Statistic

Another test statistic designed for continuous X was first proposed by Cramér in 1928 and also by Von Mises in 1931 (see Darling (10)). The statistic is defined by

$$CM = n \int_{-\infty}^{+\infty} [F_n(x) - F_0(x)]^2 \, d F_0(x)$$

$$= \frac{1}{12n} + \sum_{i=1}^{n} (U_{(i)} - \frac{2i-1}{2n})^2 .$$

It is necessary to use the limiting null distribution of CM to approximate the significance level. The exact null distribution of CM for $n = 1, 2, 3$ was examined by Marshall (30) and the agreement at $n = 3$ with the limiting distribution is remarkably close. Marshall also summarized results on sensitivity and showed that there are alternatives for which the test is biased.

Several modifications of the Cramér-Von Mises statistic have been given. Anderson and Darling (1) proposed a weighted version intended to increase sensitivity against discrepancies from F_0 in the tails. This statistic is defined by

$$WCM = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F_0(x)]^2}{F_0(x)[1 - F_0(x)]} \, d F_0(x) .$$

Another modification, introduced by Watson (40), is given by

$$WA = n \int_{-\infty}^{\infty} \{F_n(x) - F_0(x) - \int_{-\infty}^{\infty} [F_n(x) - F_0(x)] d\, F_0(x)\}^2 d\, F_0(x)$$

$$= \frac{1}{12n} - n(\overline{U} - \frac{1}{2}) + \sum_{i=1}^{n} (U_{(i)} - \frac{2i-1}{2n})^2 ,$$

where

$$\overline{U} = \frac{1}{n} \sum_{i=1}^{n} U_{(i)} .$$

Statistics Based on Spacings

The null hypothesis that $U_i = F_0(X_i)$, $i = 1, 2, \ldots, n$, are independent uniform variables on $(0, 1)$ when X is of the continuous type is equivalent to hypothesizing that the observed values of U_i, $i = 1, 2, \ldots, n$, are randomly scattered on the $(0, 1)$ interval. If the distribution departs from the null one would expect some intervals between adjacent points to be shorter and some longer than would be expected from random scatter. This suggests that a study of the relative lengths between adjacent points might be appropriate for goodness of fit.

Again let us denote the ordered transformed variables by $0 \leq U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n)} \leq 1$, and define the random interval lengths or "spacings" by

$$C_1 = U_{(1)}$$

$$C_i = U_{(i)} - U_{(i-1)}, \quad i = 2, 3, \ldots, n$$

$$C_{n+1} = 1 - U_{(n)} .$$

A number of test statistics have been proposed of the form

$$G_n = \sum_{i=1}^{n+1} g_n(C_i) .$$

Procedures based on such statistics are discussed and defended by Pyke (36) and extensively studied by Weiss (50). Examples of $g_n(\cdot)$ are $g_n(C_i) = C_i^r$, $r > 0$; $g_n(C_i) = |C_i - (n+1)^{-1}|$; $g_n(C_i) = [C_i - (n+1)^{-1}]^2$; $g_n(C_i) = \log(C_i)$; $g_n(C_i) = 1/C_i$. Note that

$$E\{C_i | U\text{'s uniform}\} = (n+1)^{-1} .$$

In each case the limiting null distribution is obtained (36).

Another important procedure has been given by Durbin (12). Durbin shows that if Y_j, $j = 1, 2, \ldots, n$, and W_r, $r = 1, 2, \ldots, n$, are defined by

$$Y_j = (n + 2 - j)(C_{(j)} - C_{(j-1)}) , \quad j = 1, 2, \ldots, n$$

and

$$W_r = \sum_{j=1}^{r} Y_j , \quad r = 1, 2, \ldots, n ,$$

where $C_{(j)}$, $j = 1, 2, \ldots, n$, are obtained by ordering the C_i, $i = 1, 2, \ldots, n$, and $C_{(0)} = 0$, then W_r, $r = 1, 2, \ldots, n$, have the same null distribution as the ordered uniform variables $U_{(i)}$, $i = 1, 2, \ldots, n$.

Durbin gives a heuristic argument that if the true distribution departs from the uniform distribution in any manner, except a change only in location, the W_r will tend to diminish. Thus any one-sided version of other goodness of fit test statistics can be applied to the W_r, for example

$$D = \max_{r=1,2,\ldots,n} [r/n - W_r] ,$$

a one-sided version of the Kolmogorov-Smirnov statistic. Computations suggest that this has good sensitivity against alternatives with the mean and variance equal to those of the hypothesized distribution (42).

Composite Goodness of Fit Hypothesis

The use of the chi-squared statistic in the composite case was first studied by Fisher (14). The approach was to use as a measure of discrepancy between the sample and hypothesized class $\{F_0(\cdot\,;\theta), \theta \in \Omega\}$

$$FCS = \min_{\theta} \{CS\} ,$$

where CS is defined in the previous section. If θ is composed of m real parameters, then under quite general conditions, FCS is approximately distributed as a chi-squared variable with $n-1-m$ (n = sample size) degrees of freedom when the null hypothesis is true.

The minimization with respect to θ can be cumbersome and several modifications have been proposed (33). The most appealing modification from a practical point of view is to replace θ with its maximum likelihood estimator. This tends to inflate FCS beyond values predicted by the chi-squared distribution leading to some unwarranted small significance levels. However, Chernoff and Lehmann (7) and also Watson (49) have shown that no serious distortion will result if the number of intervals is ten or more.

When θ is the pair (μ, σ^2), μ the mean and σ^2 the variance, $F_0(x\,;\theta)$ may be written $F_0[(x-\mu)/\sigma]$ where $F_0(\cdot)$ is completely specified. An interesting approach in this case is to transform the composite null hypothesis into an equivalent simple hypothesis. This makes it possible to test the composite hypothesis with test statistics described in the previous section and those developed in this study.

Durbin (12) proposed the following transformation to eliminate the mean and variance for the null hypothesis of normality. Let $x_1, x_2, \ldots, x_n$ be a random sample from the population of interest, and let $y_1, y_2, \ldots, y_n$ be a random sample generated synthetically from a normal population with mean zero and variance one. Let $\bar{x}$ and s_X^2 denote the sample mean and variance, respectively, of the x's, and $\bar{y}$ and s_Y^2 the sample mean and variance of the y's. Define z_i, $i = 1, 2, \ldots, n$, by

$$z_i = \frac{s_Y(x_i - \bar{x})}{s_X} + \bar{y} .$$

Durbin proves that the random variables $Z_1, Z_2, \ldots, Z_n$ are independent standard normal variables. Therefore,the composite hypothesis concerning $x_1, x_2, \ldots, x_n$ can be tested as a simple hypothesis concerning $z_1, z_2, \ldots, z_n$. The price paid for the elimination of nuisance parameters by this method is that an element of randomization is introduced in the analysis of the data.

Sarkadi (39) gave a similar transformation defined by

$$Y_i = \frac{(X_i - \bar{X}'')}{S''} \; \psi \left[\frac{|X_{n-1} - X_n| \sqrt{n-2}}{\sqrt{2}\, S''} \right] , \quad i = 1, 2, \ldots, n-2 ,$$

where

$$\bar{X}'' = \frac{\sum_{i=1}^{n} X_i + \sqrt{\frac{n}{2}}\,(X_{n-1} + X_n)}{n + \sqrt{2n}} ,$$

$$S'' = \sqrt{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\,(\sum_{i=1}^{n} X_i)^2 - \frac{1}{2}\,(X_{n-1} - X_n)^2} ,$$

and $\psi(\cdot)$ is a monotone decreasing function. The restriction that the Y_i be independent and normally distributed determines the functions $\psi(\cdot)$ completely. This transformation has the desirable property that only random samples from a normal lead to independent standard normal variables; that is, each Y has the standard normal distribution if and only if the X's have the same normal distribution. This is commonly called a "characterization" of the normal distribution. This property is important from the point of view of the biasedness of the test. This transformation also has the property that it maximizes the minimum correlation between Y_i and X_i among transformations of this general type. Sarkadi argues that this is important in that the transformed variables give a best representation of the original variables.

Notice that Sarkadi's method decreases the number of variables by two, while Durbin's method gives the same number of transformed variables as that of the original variables. Durbin's method has the disadvantage that random numbers are used in the analysis of the data which permits different investigators to draw different conclusions from the same set of data.

Other characterizations of the normal distribution as well as characterizations of other parametric families are given in the literature. The basic theory and some important results are given in a paper by Prohorov (36).

Another important approach to testing for normality is given by Shapiro and Wilk (41). They give a statistic for which under normality the numerator and denominator are both up to a constant, estimating σ^2. Let $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ denote the order statistics arising from a random sample of size n from a standard normal population. Let $\vec{m}' = (m_1, m_2, \ldots, m_n)$ denote the transpose of the vector of expected values of standard normal order statistics, and $V = (v_{ij})$ be the corresponding $n \times n$ covariance matrix. If $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$ are order statistics arising from a random sample from a normal population with μ and σ^2 unknown, then $Y_{(i)}$ may be expressed as

$$Y_{(i)} = \mu + \sigma X_{(i)} = \mu + \sigma m_i + \sigma(X_{(i)} - m_i)$$
$$= \mu + \sigma m_i + e_i \,, \quad i = 1, 2, \ldots, n \,.$$

where

$$E(e_i) = 0 \,, \quad i = 1, 2, \ldots, n \,.$$

and

$$\mathrm{Cov}(\vec{e}) = \sigma^2 V,$$

It follows from generalized least-squares that the best linear unbiased estimates of μ and σ^2 are

$$\hat{\mu} = \overline{y}$$

and

$$\hat{\sigma} = \frac{\vec{m}' V^{-1} \vec{Y}}{\vec{m}' V^{-1} \vec{m}}$$

$$\vec{Y} = \begin{pmatrix} Y_{(1)} \\ Y_{(2)} \\ \cdot \\ \cdot \\ \cdot \\ Y_{(n)} \end{pmatrix}.$$

The test statistic for normality is then defined by

$$W = \frac{R^4 \hat{\sigma}^2}{c^2 s^2} = \frac{(\vec{a}'\vec{Y})^2}{s^2} = \frac{\left(\sum_{i=1}^{n} a_i Y_{(i)}\right)^2}{\sum_{i=1}^{n} (Y_{(i)} - \bar{Y})^2},$$

where

$$s^2 = \sum_{i=1}^{n} (Y_{(i)} - \bar{Y})^2 ,$$

$$R^2 = \vec{m}' V^{-1} \vec{m} ,$$

$$c^2 = \vec{m}' V^{-1} V^{-1} \vec{m} ,$$

$$a' = (a_1, \ldots, a_n) = \frac{\vec{m}' V^{-1}}{(\vec{m}' V^{-1} V^{-1} \vec{m})^{1/2}} .$$

If the Y's are normal order statistics then the numerator and denominator are both, up to a constant, estimating σ^2. The ratio of estimates of σ^2 is multiplied by R^4/c^2 so that the linear coefficients of the $Y_{(i)}$ are normalized. Heuristic considerations augmented by extensive synthetic sampling using a wide range of populations suggest that the mean of W for non-normal populations tends to shift to the left of that for the null case. Thus, small values of W are taken to be consistent with the alternative.

It is easily shown that the distribution of W does not depend on the values of μ and σ^2, but the exact null distribution of W is not known for $n > 4$. In fact the elements of V are not known for $n > 20$. The authors first approximate the elements of V and then approximate the null distribution of W by synthetic sampling.

Combining Independent Significance Levels

The most widely used method of combining independent significance levels is the so-called omnibus test of R. A. Fisher (15) which is given by

$$F = -2 \log \left[\prod_{i=1}^{p} L_i \right] = -2 \left[\sum_{i=1}^{p} \log (L_i) \right] .$$

The null distribution of F is a chi-square with 2p degrees of freedom. Small values of the levels are consistent with H''_A, so large values of F are consistent with H''_A. The combined significance level is given by

$$\ell_F = P[\chi^2 (2p) \geq f]$$

where f is the observed value of F and $\chi^2(2p)$ is a chi-square variable with 2p degrees of freedom.

Independent of Fisher's work, E. S. Pearson (33) proposed

$$-2 \left[\sum_{i=1}^{p} \log (1 - L_i) \right]$$

as a method of combining levels. Again the statistic is distributed as a chi-squared variable under the null hypothesis, but now small values are taken to be consistent with the alternative. Others have suggested the maximum or minimum or m^{th} largest among the levels. Each of these has a beta distribution under the null hypothesis (51).

T. Liptak (27) pointed out the need, in some cases, to weight the levels differently. He proposed a statistic of the form

$$\sum_{i=1}^{p} \alpha_i \psi^{-1}(L_i)$$

where ψ^{-1} is the inverse of an arbitrary continuous distribution function and the α_i are arbitrary weights. To simplify the null distribution the obvious choice for ψ is the standard normal distribution. Under H_0'' this statistic is then distributed normally with any set of weights.

Several criteria for comparing methods of combination have been developed (4)(28). Some of these criteria are used in Chapter IV to compare methods described above with those developed in Chapter III.

CHAPTER III

A JOINT INTEGRAL TRANSFORM APPROACH

As illustrated in Chapter I, the problem of combining independent levels of significance of test statistics of the continuous type can be reduced to a one-sided goodness of fit problem. That is, if T_i, $i = 1, 2, \ldots, p$, are independent test statistics of the continuous type with levels L_i, $i = 1, 2, \ldots, p$, respectively, then under the combined null hypothesis the L_i are mutually independent uniform variables. Under the combined alternative one or more of the variables L_i are stochastically smaller than uniform variables. In this chapter a number of test statistics are constructed that can be used for either combining independent levels or the one-sided goodness of fit problem.

The Joint Integral Method of Combination

The method of combination given in this section is similar to Fisher's method, yet takes advantage of properties of uniform order statistics. It is hoped that this will increase sensitivity, particularly when not all levels are stochastically smaller than uniform variables under the combined alternative; i.e., some of the levels are uniform under both the null and alternative hypotheses.

Consider an observed value of Fisher's statistic. For observed significance levels ℓ_i, $i = 1, 2, \ldots, p$, $0 < \ell < 1$, an observed value of F is given by

$$f = -2\log\left[\prod_{i=1}^{p} \ell_i\right] = -2\log\left[\prod_{i=1}^{p} P_0\{L_i \leq \ell_i\}\right]$$

$$= -2\log\left[G_0(\ell_1, \ell_2, \ldots, \ell_p)\right]$$

where $P_0[\cdot]$ denotes probability under the combined null and $G_0(\cdot)$ the joint distribution function of the levels under the combined null. Thus the joint null distribution function provides the measure of departure from uniform.

The approach here is to use as a measure of departure from uniform variables, the joint null distribution function of the ordered observed levels. Again let ℓ_i, $i = 1, 2, \ldots, p$, denote the observed significance levels, then an observed value of the test statistic T, henceforth referred to as the joint integral transform method of combination, will be defined by

$$t = K_0(\ell_{(1)}, \ell_{(2)}, \ldots, \ell_{(p)})$$

where $0 \leq \ell_{(1)} \leq \ell_{(2)} \leq \cdots \leq \ell_{(p)} \leq 1$ are the ordered observed levels and $K_0(\cdot)$ is the joint distribution function of p uniform order statistics. The value of t can be obtained by evaluating the multiple integral

$$t = p! \int_0^{\ell_{(1)}} \int_{u_1}^{\ell_{(2)}} \cdots \int_{u_{p-2}}^{\ell_{(p-1)}} \int_{u_{p-1}}^{\ell_{(p)}} du_p \, du_{p-1} \cdots du_2 \, du_1 .$$

In this and subsequent chapters, density and distribution functions of uniform order statistics, order statistics arising of independent uniform variables, will be given without proof. For a complete discussion of order statistics, the reader is referred to almost any probability or mathematical statistics textbook; for example, see Fisz (16).

To obtain a more direct expression for t consider the following theorem and proof given by Suzuki (45).

<u>Theorem 3.1</u>: If $0 \leq U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n)} \leq 1$ are uniform order statistics and a_i, $i = 1, 2, \ldots, n$, are real numbers such that

$$0 \leq a_1 \leq a_2 \leq \cdots \leq a_n \leq 1$$

then

$$P[U_{(i)} \geq a_i,\ i = 1, 2, \ldots, n] = \sum_{k=0}^{n} \binom{n}{k} q_k ,$$

where

$$q_0 = 1$$

$$q_k = - \sum_{i=0}^{k-1} \binom{k}{i} a_k^{k-i} q_i ,\quad k = 1, 2, \ldots, n .$$

Proof: Denote $P[U_{(i)} \geq a_i,\ i = 1, 2, \ldots, n]$ by $p_n(a_1, a_2, \ldots, a_n)$. For $n = 1$,

$$p_1(a_1) = \int_{a_1}^{1} du = 1 - a_1 .$$

For $n = 2, 3, \ldots$

$$p_n(a_1, \ldots, a_n) = n! \int_{a_n}^{1} \int_{a_{n-1}}^{u_n} \cdots \int_{a_2}^{u_3} \int_{a_1}^{u_2} du_1 du_2 \ldots du_{n-1} du_n$$

$$= n! \int_{\Gamma} \cdots \int du_1 du_2 \ldots du_n ,$$

where

$$\Gamma = \{(u_1, \ldots, u_n) \mid u_i \geq a_i, \; i = 1, 2, \ldots, n \; ; \; 0 \leq u_1 \leq u_2 \leq \ldots \leq u_n \leq 1\} \; .$$

Consider the change of variables

$$\begin{aligned} y_1 &= u_1/u_n \\ y_n &= u_2/u_n \\ &\vdots \\ y_{n-1} &= u_{n-1}/u_n \\ y_n &= u_n \; . \end{aligned}$$

The region Γ corresponds in a one-to-one fashion to the region Γ^* given by

$$\Gamma^* = \{(y_1, \ldots, y_n) \mid y_i \geq a_i/y_n \; ; \; 0 \leq y_1 \leq \ldots \leq y_{n-1} \leq 1 \; ; \; a_n \leq y_n \leq 1\} \; .$$

The Jacobian is

$$J = y_n^{n-1} \; ,$$

so

$$p_n(a_1, \ldots, a_n) = n! \int_{a_n}^{1} \left\{ \int_{a_{n-1}/y_n}^{1} \cdots \int_{a_1/y_n}^{y_2} dy_1 \ldots dy_{n-1} \right\} y_n^{n-1} \, dy_n$$

$$= n \int_{a_n}^{1} p_{n-1}(a_1/y_n, \ldots, a_{n-1}/y_n) \, y_n^{n-1} \, dy_n \; .$$

Consider $q_k(a_1, a_2, \ldots, a_k)$ defined by

$$q_0 = 1$$

$$q_k(a_1, \ldots, a_k) = - \sum_{i=0}^{k-1} \binom{k}{i} a_k^{k-i} q_i , \quad k = 1, 2, \ldots, n .$$

Note that q_k is a homogeneous polynomial of degree k. Now

$$p_1(a_1) = 1 - a_1 = q_0 + q_1(a_1) .$$

Suppose

$$p_{n-1}(a_1, \ldots, a_{n-1}) = \sum_{k=0}^{n-1} \binom{n-1}{k} q_k ,$$

then

$$p_n(a_1, \ldots, a_n)$$

$$= n \int_{a_n}^{1} p_{n-1}(a_1/y, \ldots, a_{n-1}/y)\, y^{n-1} dy$$

$$= n \int_{a_b}^{1} \left\{ \sum_{k=0}^{n-1} \binom{n-1}{k} q_k(a_1/y, \ldots, a_k/y) \right\} y^{n-1} dy$$

$$= n \left\{ \sum_{k=0}^{n-1} \binom{n-1}{k} \int_{a_n}^{1} q_k(a_1/y, \ldots, a_k/y)\, y^{n-1} dy \right\}$$

$$= n \left\{ \sum_{k=0}^{n-1} \binom{n-1}{k} \int_{a_n}^{1} q_k(a_1, a_2, \ldots, a_k)\, y^{n-k-1} dy \right\}$$

$$= n \left\{ \sum_{k=0}^{n-1} \binom{n-1}{k} q_k(a_1, \ldots, a_k) \left[1 - a_n^{n-k} \right] / (n-k) \right\}$$

$$= \sum_{k=0}^{n-1} \frac{n}{n-k} \binom{n-1}{k} q_k(a_1, \ldots, a_k) - \sum_{k=0}^{n-1} \frac{n}{n-k} \binom{n-1}{k} a_n^{n-k} q_k(a_1, \ldots, a_k)$$

$$= \sum_{k=0}^{n-1} \binom{n}{k} q_k + q_n$$

$$= \sum_{k=0}^{n} \binom{n}{k} q_k ,$$

and the theorem is established.

Consider again an observed value of T,

$$t = K_0(\ell_{(1)}, \ldots, \ell_{(p)}) = P_0[L_{(1)} \leq \ell_{(1)}, \ldots, L_{(p)} \leq \ell_{(p)}]$$
$$= P_0[1 - L_{(1)} \geq 1 - \ell_{(1)}, \ldots, 1 - L_{(p)} \geq 1 - \ell_{(p)}].$$

Suppose we define

$$U_{(i)} = 1 - L_{(p-i+1)}, \; i = 1, 2, \ldots, p$$

and

$$a_i = 1 - \ell_{(p-i+1)}, \; i = 1, 2, \ldots, p.$$

Then $U_{(i)}$, $i = 1, 2, \ldots, p$, are uniform order statistics under the null hypothesis and the real numbers a_i, $i = 1, 2, \ldots, p$, satisfy

$$0 \leq a_1 \leq a_2 \leq \ldots \leq a_p \leq 1 .$$

Thus

$$t = \sum_{k=0}^{p} \binom{p}{k} q_k$$

where

$$q_0 = 1$$

$$q_k = - \sum_{i=0}^{k-1} \binom{k}{i} (1 - \ell_{(p-k+1)})^{k-i} q_i , \; k = 1, 2, \ldots, p .$$

For example if $p = 4$, then

$$q_0 = 1 ,$$

$$q_1 = -(1 - \ell_{(4)}) ,$$

$$q_2 = -(1 - \ell_{(3)})^2 - 2(1 - \ell_{(3)}) q_1 ,$$

$$q_3 = -(1 - \ell_{(2)})^3 - 3(1 - \ell_{(2)})^2 q_1 - 3(1 - \ell_{(2)}) q_2 ,$$

$$q_4 = -(1 - \ell_{(1)})^4 - 4(1 - \ell_{(1)})^3 q_1 - 6(1 - \ell_{(1)})^2 q_2 - 4(1 - \ell_{(1)}) q_3 ,$$

$$t = 1 + 4q_1 + 6q_2 + 4q_3 + q_4 .$$

Clearly small values of T should be taken as consistent with the alternative, however, the null distribution function of

$$T = \sum_{k=0}^{p} \binom{p}{k} Q_k(L_{(p)}, L_{(p-1)}, \ldots, L_{(p-k+1)})$$

is needed to compute the combined significance level.

Theorem 3..2: The null distribution of T for $p = 2$ is given by

$$F_2(t) = \begin{cases} 0 , & t \leq 0 \\ 1 - (1 - t)^{1/2} + t \log \left[\dfrac{t^{1/2}}{1 - (1 - t)^{1/2}} \right] , & 0 < t < 1 \\ 1 , & t \geq 1 . \end{cases}$$

Proof: In terms of $L_{(1)}$ and $L_{(2)}$,

$$T = 1 + 2Q_1 + Q_2 = 1 - 2(1 - L_{(2)}) - (1 - L_{(1)})^2 + 2(1 - L_{(1)})(1 - L_{(2)})$$

$$= L_{(1)}(2 L_{(2)} - L_{(1)}) .$$

To simplify notation let $X = L_{(1)}$ and $Y = L_{(2)}$, then the null density function of (X, Y) is

$$f_0(x, y) = 2, \quad 0 \le x \le y \le 1$$
$$= 0, \quad \text{otherwise} .$$

For t in $(0, 1)$,

$$P_0[T \le t] = P_0[X(2Y - X) \le t]$$
$$= \underset{X}{E}_0 \{ P_0[Y < t/2X + X/2 \,|\, X] \}$$

Now

$$P_0[Y \le t/2x + x/2 \,|\, X = x] = 0$$

if

$$t/2x + x/2 < x$$

or, equivalently, if

$$x > t^{1/2} ;$$

$$P[Y \le t/2x + x/2 \,|\, X = x] = \int_x^{t/2x + x/2} f_0(y|x) dy$$

if

$$x \le t/2x + x/2 \le 1$$

or, equivalently, if

$$1 - (1 - t)^{1/2} \le x \le t^{1/2} ;$$

and

$$P[Y \le t/2x + x/2 \,|\, X = x] = 1$$

if $t/2x + x/2 > 1$ or, equivalently, if $0 \leq x < 1 - (1-t)^{1/2}$. Since

$$f_0(y|x) = \frac{1}{(1-x)}, \quad x \leq y \leq 1$$
$$= 0, \quad \text{otherwise},$$

and

$$f_0(x) = 2(1-x), \quad 0 \leq x \leq 1$$
$$= 0, \quad \text{otherwise},$$

$$P_0[T_2 \leq t] = \int_0^{1-(1-t)^{1/2}} 2(1-x)dx + \int_{1-(1-t)^{1/2}}^{t^{1/2}} [t/x - x]dx$$

$$= 1 - (1-t)^{1/2} + t \log\left[\frac{t^{1/2}}{1-(1-t)^{1/2}}\right]$$

and the theorem is proved.

It was not possible to obtain an explicit form for the null distribution of T for general p. This necessitated consideration of an approximation. The first attempt was to obtain expressions for moments, however this has been possible only for the first moment.

Theorem 3.3: The mean of T, under the null hypothesis, is $(p+1)^{-1}$.

Proof: Again let $0 \leq L_{(1)} \leq L_{(2)} \leq \ldots \leq L_{(p)} \leq 1$ denote the ordered significance levels and define X_k, $k = 1, 2, \ldots, p$, by

$$X_k = 1 - L_{(p+1-k)}, \quad k = 1, 2, \ldots, p.$$

Then $0 \leq X_1 \leq X_2 \leq \ldots \leq X_p \leq 1$ are distributed like p uniform order statistics under the null hypothesis and

$$T = \sum_{k=0}^{p} \binom{p}{k} Q_k(X_1, \ldots, X_k) ,$$

where

$$Q_0 = 1$$
$$Q_k = - \sum_{i=0}^{k-1} \binom{k}{i} X_k^{k-i} Q_i(X_1, \ldots, X_i) , \; k = 1, 2, \ldots, p .$$

For

$$k = 1 , Q_1 = -X_1$$
$$k = 2 , Q_2 = -X_2^2 - 2X_2 Q_1$$

and for $k = 3, 4, \ldots$

$$Q_k = -X_k^k - k X_k^{k-1} Q_1 - \sum_{i=2}^{k-1} \binom{k}{i} X_k^{k-i} Q_i .$$

The conditional density of X_1 given $X_k, X_{k+1}, \ldots, X_p$ is

$$f(x_1 | x_k, \ldots, x_p) = \frac{(k-1)(x_k - x_1)^{k-2}}{x_k^{k-1}} , \; 0 \le x_1 \le x_k \le \ldots \le x_p$$
$$= 0 , \text{ otherwise}$$

so

$$E(X_1 | X_k, X_{k+1}, \ldots, X_p) = \frac{(k-1)}{X_k^{k-1}} \int_0^{X_k} t(X_k - t)^{k-2} \, dt = \frac{X_k}{k} .$$

Thus

$$E(Q_2 | X_2) = E\{-X_2^2 - 2X_2 Q_1 | X_2\} = -X_2^2 - 2X_2 E(-X_1 | X_2)$$
$$= -X_2^2 - 2X_2 (-\tfrac{1}{2} X_2) = 0$$

Assume that for $k > 2$,

$$E(Q_i | X_i) = 0, \quad i = 2, 3, \ldots, k-1,$$

then

$$E(Q_k | X_k) = E\{-X_k^k - kX_k^{k-1} Q_1 - \sum_{i=2}^{k-1} \binom{k}{i} X_k^{k-i} Q_i | X_k\}$$

$$= -X_k^k - kX_k^{k-1} E\{-X_1 | X_k\} - \sum_{i=2}^{k-1} \binom{k}{i} X_k^{k-i} \underset{i|k}{E} \{E(Q_i | X_i, X_k)\}$$

where $\underset{i|k}{E}$ denotes the expectation with respect to X_i given X_k. Now since $E(Q_i | X_i, X_k) = E(Q_i | X_i) = 0, \; i = 1, 2, \ldots, k-1$,

$$E(Q_k | X_k) = -X_k^k - kX_k^{k-1}\left(-\frac{1}{k} X_k\right) = 0,$$

so

$$E(Q_k) = \underset{k}{E} \{E(Q_k | X_k)\} = 0, \quad k = 2, 3, \ldots, p.$$

Now

$$E(T) = E\{1 + pQ_1 + \sum_{k=2}^{p} \binom{p}{k} Q_k\}$$

$$= 1 + pE(Q_1) = 1 - p^2 \int_0^1 t(1-t)^{p-1} dt$$

$$= 1 - \frac{p^2}{p(p+1)} = \frac{1}{p+1}.$$

It has not been possible to obtain explicit expressions for higher moments. Hence, techniques for approximating distribution functions based on moments cannot be used (22). In this circumstance it seems both appropriate and efficient to employ synthetic sampling to obtain an approximation for the null distribution of T. Accordingly, for each $p = 2, 3, \ldots, 9$, three thousand uniform samples were generated employing the IBM subroutine RANDU (46). Then for each p, 3000 values of $(p+1)T$ were computed and empirical percentage points determined. These percentage points are given in Table I. Some check of the accuracy of the approximation is provided by the exact percentage points for $p = 2$.

A summary of the steps necessary for combining independent levels by this method might be appropriate. To apply the joint integral transform method, given p observed levels $\ell_1, \ell_2, \ldots, \ell_p$, one proceeds as follows:

(i) Order the levels to obtain $0 \leq \ell_{(1)} \leq \ell_{(2)} \leq \cdots \leq \ell_{(p)} \leq 1$.

(ii) Compute $x_i = 1 - \ell_{(p+1-i)}$, $i = 1, 2, \ldots, p$.

(iii) Compute

$$q_k = - \sum_{i=0}^{k-1} \binom{k}{i} x_k^{k-i} q_i \,, \quad k = 1, 2, \ldots, p \; ; \; q_0 = 1 \,,$$

and

$$(p+1)t = (p+1) \sum_{k=0}^{p} \binom{p}{k} q_k \,.$$

(iv) Small values are taken to be consistent with the alternative, i.e., indicate nonuniform. The approximate combined significance level can be obtained from Table I.

TABLE I

EMPIRICAL PERCENTAGE POINTS

Number of Levels	Probability of Smaller Value								
	.01	.05	.10	.15	.20	.25	.50	.75	.90
2, exact	.0071	.0457	.105	.171	.247	.328	.836	1.557	2.187
2, approx.	.0070	.0448	.108	.176	.249	.329	.831	1.552	2.200
3	.0035	.0267	.065	.116	.179	.244	.745	1.535	2.364
4	.0027	.0211	.052	.088	.141	.203	.656	1.520	2.491
5	.0010	.0139	.039	.073	.116	.167	.617	1.510	2.628
6	.0008	.0103	.029	.060	.100	.150	.538	1.494	2.704
7	.0004	.0085	.024	.048	.087	.131	.533	1.449	2.802
8	.0004	.0071	.021	.044	.078	.120	.488	1.406	2.817
9	.0003	.0059	.019	.040	.073	.111	.442	1.334	2.813

Conditional Integral Transform Methods

Consider again an observed value of the joint ingegral transform statistic:

$$t = P_0[L_{(1)} \leq \ell_{(1)}, \ldots, L_{(p)} \leq \ell_{(p)}]$$

$$= \left\{ \sum_{k=1}^{p-1} P_0[L_{(k)} \leq \ell_{(k)} \mid L_{(k+1)} \leq \ell_{(k+1)}, \ldots, L_{(p)} \leq \ell_{(p)}] \right\} P_0[L_{(p)} \leq \ell_{(p)}]$$

Suppose we define a statistic, say W_1, by replacing inequalities in the conditioning statement by equalities; that is

$$w_1 = \left\{ \prod_{k=1}^{p-1} P_0[L_{(k)} \le \ell_{(k)} \mid L_{(k+1)} = \ell_{(k+1)}, \ldots, L_{(p)} = \ell_{(p)}] \right\} P_0[L_{(p)} \le \ell_{(p)}]$$

Theorem 3.4: The test statistic W_1 is the Fisher statistic.

Proof: The conditional density of $L_{(k)}$ given $L_{(k+1)}, \ldots, L_{(p)}$ is

$$f_{k|k+1,\ldots,n}(\ell_{(k)} \mid \ell_{(k+1)}, \ldots, \ell_{(p)}) = \frac{k\,\ell_{(k)}^{k-1}}{\ell_{(k+1)}}, \quad 0 \le \ell_{(k)} \le \ell_{(k+1)},$$

$$= 0, \quad \text{otherwise}.$$

and the conditional distribution function is

$$P_0[L_{(k)} \le \ell_{(k)} \mid L_{(k+1)} = \ell_{(k+1)}, \ldots, L_{(p)} = \ell_{(p)}] = \ell_{(k)}^{k} / \ell_{(k+1)}^{k}, \quad k = 1, 2, \ldots, n-1,$$

$$P_0[L_{(p)} \le \ell_{(p)}] = \ell_{(p)}^{p}.$$

An observed value of W_1 is then

$$w_1 = \frac{\ell_{(1)}}{\ell_{(2)}} \frac{\ell_{(2)}^2}{\ell_{(3)}^2} \cdots \frac{\ell_{(p-1)}^{p-1}}{\ell_{(p)}^{p-1}} \ell_{(p)}^{p} = \prod_{k=1}^{p} \ell_{(k)} = \prod_{k=1}^{p} \ell_k$$

and the theorem is proved.

Similarly the statistic defined by

$$w_2 = P_0[L_{(1)} \ge \ell_{(1)}] \left\{ \prod_{k=2}^{p} P_0[L_{(k)} \ge \ell_{(k)} \mid L_{(1)} = \ell_{(1)}, \ldots, L_{(k-1)} = \ell_{(k-1)}] \right\}$$

is equivalent to the Pearson statistic if large values are taken to be consistent with the alternative. Of course, neither of these offer new methods of combination; however, the approach can be used to generate a number of statistics with known null distribution functions. The following theorem will provide both a method for constructing test statistics and the basis for finding the null distribution of such statistics.

Theorem 3.5: If the conditional distribution function, say $F_{X|Y}(x|y)$, of the random variable X (a scalar) given the random variable Y (either scalar or vector) is such that for each value y in some interval (c, d),

(1) there exists an $a = a(y)$ and $b = b(y)$ such that $F_{X|Y}(a|y) = 0$ and $F_{X|Y}(b|y) = 1$,

(2) $F_{X|Y}(x|y)$ is a continuous and strictly increasing function of x for $a < x < b$, then

(i) $U = F_{X|Y}(X|Y)$ is distributed uniformly on the interval $(0, 1)$,

(ii) U and Y are independent.

Proof: Consider the conditional distribution of U given $Y = y \in (c, d)$; let $0 < u < 1$, then

$$G_{U|Y}(u|y) = P[U \leq u | Y = y] = P[F_{X|Y}(X|Y) < u | Y = y].$$

Because of (2) there is a point x_0 satisfying $F_{X|Y}(x_0|y) = u$ and $F_{X|Y}(x|y) \leq u$ for $x \leq x_0$. Then

$$G_{U|Y}(u|y) = P[X \leq x_0 | Y = y] = F_{X|Y}(x_0|y) = u .$$

If $u \geq 1$, then $G_{U|Y}(u|y) = 1$ and if $u \leq 0$, then $G_{U|Y}(u|y) = 0$. Thus the conditional distribution function of U given $Y = y$ is

$$G_{U|Y}(u|y) = \begin{cases} 0 , & u \leq 0 , \\ u , & 0 < u < 1 , \\ 1 , & u \geq 1 . \end{cases}$$

Because $G_{U|Y}(u|y)$ does not depend on the value of Y, U and Y are independent and U is uniformly distributed.

The preceding theorem will be used to obtain mutually independent uniform variables. Suppose $Y_1, \ldots, Y_m$ have joint distribution function $F(y_1, \ldots, y_m)$ and define

$$Z_1 = F_{Y_1|Y_2 \cdots Y_m}(Y_1|Y_2, \ldots, Y_m) ,$$

where $F_{Y_1|Y_2 \cdots Y_m}(\cdot|\cdot)$ denotes the conditional distribution function of Y_1 given $Y_2, Y_3, \ldots, Y_m$. If this conditional distribution satisfies the conditions of Theorem 3.5, then Z_1 is uniformly distributed and independent of $(Y_2, Y_3, \ldots, Y_m)$. Now define Z_2 by

$$Z_2 = F_{Y_2|Y_3 \cdots Y_m}(Y_2|Y_3, \ldots, Y_m) .$$

Again if $F_{Y_2|Y_3 \cdots Y_m}(\cdot|\cdot)$ satisfies the conditions of Theorem 3.5, Z_2 is a uniform variable and Z_2 is independent of $(Y_3, \ldots, Y_m)$, and the three random variables Z_1, Z_2 and $(Y_3, \ldots, Y_m)$ are mutually independent since

$$P[Z_1 \le z_1,\ Z_2 \le z_2,\ Y_3 \le y_3, \ldots, Y_m \le y_m]$$

$$= P[Z_1 \le z_1] \cdot P[Z_2 \le z_2,\ Y_3 \le y_3, \ldots, Y_m \le y_m]$$

$$= P[Z_1 \le z_1] \cdot P[Z_2 \le z_2] \cdot P[Y_3 \le y_3, \ldots, Y_m \le y_m]$$

for all z_1, z_2, and $(y_3, \ldots, y_m)$. Continuing this process, m mutually independent uniform variables can be obtained.

Of course, the statistics W_1 and W_2 were constructed using this technique and it was clear that, under the null hypothesis, each was the product of p independent uniform variables. Consider now reversing the inequalities in W_1 ; i.e., define an observed value of a test statistic by

$$w_3 = \left\{ \prod_{k=1}^{p-1} P_0[L_{(k)} \ge \ell_{(k)} \mid L_{(k+1)} = \ell_{(k+1)}, \ldots, L_{(p)} = \ell_{(p)}] \right\} P_0[L_{(p)} \ge \ell_{(p)}]$$

$$= \left\{ \prod_{k=1}^{p-1} (1 - P_0[L_{(k)} \le \ell_{(k)} \mid L_{(k+1)} = \ell_{(k+1)}, \ldots, L_{(p)} = \ell_{(p)}] \right\} \left\{ 1 - P_0[L_{(p)} \le \ell_{(p)}] \right\}$$

$$= \prod_{k=1}^{p-1} [1 - \ell_{(k)}^k / \ell_{(k+1)}^k] [1 - \ell_{(p)}^p] .$$

Large values of W_3 will be taken to be consistent with the combined alternative. Similarly, reversing the inequalities in W_2, define W_4 by

$$w_4 = P_0[L_{(1)} \le \ell_{(1)}] \left\{ \prod_{k=2}^{p} P_0[L_{(k)} \le \ell_{(k)} \mid L_{(1)} = \ell_{(1)}, \ldots, L_{(k-1)} = \ell_{(k-1)}] \right\}$$

$$= [1 - (1 - \ell_{(1)})^p] \prod_{k=2}^{p} \left[1 - \frac{(1 - \ell_{(k)})^{p-k+1}}{(1 - \ell_{(k-1)})^{p-k+1}} \right]$$

Small values are taken consistent with the combined alternative.

Since Theorem 3.5 applies in both cases, both W_3 and W_4 are distributed as the product of p mutually independent uniform variables under the null hypothesis. Thus the observed combined significance levels are

$$\ell_3 = P[\chi^2(2n) \geq -2\ell n\, t_3],$$

and

$$\ell_4 = P[\chi^2(2n) \leq -2\ell n\, t_4]$$

where w_3 and w_4 represent observed values of W_3 and W_4, respectively, and $\chi^2(2n)$ indicates a chi-squared random variable with $2n$ degrees of freedom.

In all statistics defined above, the conditioning process began either with the first or last order statistic. Theorem 3.5 will apply regardless of where conditioning begins and the process will yield mutually independent uniform variables if once an order statistic appears left of the conditioning statement, it does not appear later in the process. For example, let $p = 4$ and define Z_1 by

$$z_1 = P_0[L_{(3)} \leq \ell_{(3)} | L_{(1)} = \ell_{(1)}, L_{(2)} = \ell_{(2)}; L_{(4)} = \ell_{(4)}]$$

$$= \frac{1}{\ell_{(4)} - \ell_{(2)}} \int_{\ell_{(2)}}^{\ell_{(3)}} dt = \frac{\ell_{(3)} - \ell_{(2)}}{\ell_{(4)} - \ell_{(2)}},$$

define Z_2 by

$$z_2 = P_0[L_{(2)} \leq \ell_{(2)} | L_{(1)} = \ell_{(1)},\ L_{(4)} = \ell_{(4)}]$$

$$= \frac{2}{[\ell_{(4)} - \ell_{(1)}]^2} \int_{\ell_{(1)}}^{\ell_{(2)}} [\ell_{(4)} - t]dt$$

$$= \frac{1}{[\ell_{(4)} - \ell_{(1)}]^2} \left\{ [\ell_{(4)} - \ell_{(1)}]^2 - [\ell_{(4)} - \ell_{(2)}]^2 \right\}$$

$$= 1 - \left[\frac{\ell_{(4)} - \ell_{(2)}}{\ell_{(4)} - \ell_{(1)}} \right]^2$$

define Z_3 by

$$z_3 = P_0[L_{(4)} \leq \ell_{(4)} | L_{(1)} = \ell_{(1)}]$$

$$= \frac{3}{[1 - \ell_{(1)}]^3} \int_{\ell_{(1)}}^{\ell_{(4)}} (t - \ell_{(1)})^2 dt$$

$$= \frac{[\ell_{(4)} - \ell_{(1)}]^3}{[1 - \ell_{(1)}]^3} ,$$

and define Z_4 by

$$z_4 = P_0[L_{(1)} \leq \ell_{(1)}] = 1 - [1 - \ell_{(1)}]^4 .$$

Now if we define W_5 as

$$W_5 = Z_1 \cdot Z_2 \cdot Z_3 \cdot Z_4 ,$$

then, under the combined null hypothesis, $-2\ell n\ W_5$ is again distributed as a chi-squared with 8 degrees of freedom. If small values of W_5 are taken to be consistent with the combined alternative, the significance level is

$$\ell_5 = P[\chi^2(2n) > -2\ell n\, t_5].$$

CHAPTER IV

PROPERTIES OF METHODS OF COMBINATION

In this chapter properties of the various methods of combination will be studied with the hope that comparisons will provide a choice among the methods. Birnbaum (4) made a study of sensitivity of four methods of combination, specifically, Fisher's method, Pearson's method, the minimum of the p levels, and the maximum of the p levels. Birnbaum assumed that each of the original samples, on which a test statistic is based, has density function of the Fisher-Koopman-Pitman-Darmois (FKPD) form, which is

$$f(x, \theta) = c(\theta)\, a(\theta)^{t(x)}\, b(x)$$

where θ is the parameter and $a, b, c,$ and t denote arbitrary functions. Among the four methods considered, Birnbaum concluded that to combine independent tests on FKPD form distributions (these include most distributions commonly occuring in applied statistics) one should choose between Fisher's method and the minimum of the individual levels.

Littell (29) compared three of these methods (all except Pearson's method) based on an asymptotic criteria. Littell concluded that of these three methods, Fisher's method is generally preferable based on this criteria. These results will be extended in the second section of this chapter. The properties of unbiasedness and consistency of methods of combination are considered in the following section.

Unbiasedness and Consistency

Birnbaum (4) introduced the idea of monotone methods of combination defined below:

Definition 4.1: Suppose $W = W(L_1, L_2, \ldots, L_p)$ is a method of combination with small [large] values consistent with the combined alternative. Then W is a monotone method of combination if

$$W(\ell_1, \ell_2, \ldots, \ell_p) \leq W(\ell_1^*, \ell_2^*, \ldots, \ell_p^*)$$

$$[W(\ell_1, \ell_2, \ldots, \ell_p) \geq W(\ell_1^*, \ell_2^*, \ldots, \ell_p^*)]$$

when $\ell_i \leq \ell_i^*$, $i = 1, 2, \ldots, p$. Birnbaum proved that the most sensitive method of combination for any particular (completely specified) alternative statisfies this condition and thus concluded that any method of combination which failed to satisfy the condition would seem unreasonable.

Clearly Fisher's method, Pearson's method, the method based on the minimum of the p levels, and that based on the maximum of the p levels are all monotone methods of combination. The joint integral method, say T, is also monotone since if $\ell_i \leq \ell_i^*$, $i = 1, 2, \ldots, p$, then the ordered values $\ell_{(i)}$ and $\ell_{(i)}^*$ also satisfy $\ell_{(i)} \leq \ell_{(i)}^*$, $i = 1, 2, \ldots, p$, and

$$t(\ell_1, \ell_2, \ldots, \ell_p) = P_0[L_{(1)} \leq \ell_{(1)}, L_{(2)} \leq \ell_{(2)}, \ldots, L_{(p)} \leq \ell_{(p)}]$$
$$\leq P_0[L_{(1)} \leq \ell^*_{(1)}, L_{(2)} \leq \ell^*_{(2)}, \ldots, L_{(p)} \leq \ell^*_{(p)}]$$
$$= t(\ell^*_1, \ell^*_2, \ldots, \ell^*_p) .$$

A relationship between monotonicity and unbiasedness (see Definition 1.4) is given in the following theorem. This theorem is similar to that proved by Chapman (6) in connection with one-sided goodness of fit statistics.

Theorem 4.1: Suppose T_i, $i = 1, 2, \ldots, p$, with significance levels L_i, $i = 1, 2, \ldots, p$, are unbiased continuous test statistics for $H_{0,i}$ versus $H_{A,i}$, $i = 1, 2, \ldots, p$, respectively. If $W(L_1, L_2, \ldots, L_p)$ is a monotone method of combination then W is unbiased for the combined hypothesis problem.

Proof: Let $H_{\theta_i}(\cdot)$, $i = 1, 2, \ldots, p$, $\theta_i \in \Omega_i$, denote the distribution functions of L_i, $i = 1, 2, \ldots, p$, respectively, and let $H^W_\theta(\cdot)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$, $\theta \in \Omega_1 \times \Omega_2 \times \ldots \times \Omega_p$, denote the distribution function of the combined significance level using W as a method of combination.

Since each of the T_i is unbiased and of the continuous type,

$$H_{\theta_i}(\ell) = \ell , \ell \in [0, 1], \ \theta_i \in \Omega_{0,i} ,$$

$$H_{\theta_i}(\ell) \geq \ell , \ \ell \in [0, 1], \ \theta_i \in \Omega_{A,i} .$$

We will not require W to be of the continuous type, so it is necessary to show that for each $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$ and each achievable

$\ell \in [0, 1]$,

$$H_\theta^W(\ell) \geq \ell .$$

Accordingly, let $\theta = (\theta_1, \ldots, \theta_p)$, $\theta_i \in \Omega_i$, and ℓ^* be an achievable level for this θ. Then there is a w_{ℓ^*} such that

$$\begin{aligned}\ell^* &= P_0[W(L_1, L_2, \ldots, L_p) \leq w_{\ell^*}] \\ &= \int \cdots_R \int \prod_{i=1}^{p} d\ell_i\end{aligned}$$

where $R = \{(\ell_1, \ell_2, \ldots, \ell_p) \mid W(\ell_1, \ell_2, \ldots, \ell_p) \leq w_{\ell^*}\}$. The distribution function of the combined level ℓ^* is given by

$$\begin{aligned}H_\theta^W(\ell^*) &= P_\theta[W(L_1, L_2, \ldots, L_p) \leq w_{\ell^*}] \\ &= \int \cdots_R \int \prod_{i=1}^{p} d\, H_{\theta_i}(\ell) .\end{aligned}$$

Make the change of variables

$$Y_i = H_{\theta_i}(L_i), \quad i = 1, 2, \ldots, p$$

then the Y_i are mutually independent uniform random variables and

$$H_\theta^W(\ell^*) = \int \cdots_S \int \prod_{i=1}^{p} d\, y_i$$

where

$$S = \{(y_1, y_2, \ldots, y_p) \mid W(H_{\theta_1}^{-1}(y_1), \ldots, H_{\theta_p}^{-1}(y_p)) \leq w_{\ell^*}\} ,$$

and

$$H_{\theta_i}^{-1}(y_i) = \inf_{0 \leq \ell \leq 1} \{\ell \mid H_{\theta_i}(\ell) = y_i\} .$$

Now since

$$H_{\theta_i}(\ell) \geq \ell , \quad \ell \in [0, 1] ,$$

we have

$$H_{\theta_i}^{-1}(\ell) \leq \ell , \quad \ell \in [0, 1] ,$$

and since W is monotone

$$W(y_1, \ldots, y_p) \geq W(H_{\theta_1}^{-1}(y_1), \ldots, H_{\theta_p}^{-1}(y_p)) .$$

So if we define S^* to be

$$S^* = \{(y_1, \ldots, y_p) \mid W(y_1, \ldots, y_p) \leq w_{\ell^*}\}$$

then $S^* \subset S$, and $S^* = R$, so

$$H_\theta^W(\ell^*) = \int \cdots_S \int \prod_{i=1}^{p} d\, y_i \geq \int \cdots_{S^*} \int \prod_{i=1}^{p} d\, y_i = \ell^*$$

and the theorem is proved.

It is also possible to obtain a relationship between monotonicity and consistency (see Definition 1.5) if some restrictions are placed on the method of combination and on the combined alternative. Specifically, assume that the method of combination is of the continuous type for continuous T_i, $i = 1, 2, \ldots, p$ and assume that the combined alternative of interest is

$$H_A^*: \theta_i \in \Omega_{A,i} \quad \text{for all} \quad i = 1, 2, \ldots, p .$$

Now let $T_i^{(n)}$, $i = 1, 2, \ldots, p$ denote the p independent test statistics where n is the sample size on which each of the individual test statistics is based. If we let $L_i^{(n)}$ denote the significance level of $T_i^{(n)}$ based on sample of size n, then if the sequence $\{T_i^{(n)}\}$ is consistent for $H_{0,i}$ versus $H_{A,i}$, the sequence $\{L_i^{(n)}\}$ converges to zero in probability $[\theta_i]$, $\theta_i \in \Omega_{A,i}$ (see Definition 1.5). For this theorem we shall denote $W(L_1^{(n)}, L_2^{(n)}, \ldots, L_p^{(n)})$ by W_n.

<u>Theorem 4.2</u>: If the sequences $\{T_i^{(n)}\}$, $i = 1, 2, \ldots, p$, are consistent for $H_{0,i}$ versus $H_{A,i}$, $i = 1, 2, \ldots, p$, respectively, and if $W(L_1^{(n)}, \ldots, L_p^{(n)})$ is of the continuous type for $\theta = (\theta_1, \theta_2, \ldots, \theta_p) \in \Omega_1 \times \Omega_2 \times \ldots \times \Omega_p$, and is monotone, then the sequence $\{W_n\}$ is consistent for H_0 versus H_A^*.

Proof: The null distribution of W_n does not change as $n \to \infty$ since $L_i^{(n)}$ is a uniform variable for $\theta_i \in \Omega_{0,i}$ regardless of the sample size n. Let $H_\theta^{(n)}(\ell)$, $\theta = (\theta_1, \ldots, \theta_p) \in \Omega_1 \times \Omega_2 \times \ldots \times \Omega_p$, be the distribution of the level of W_n; then it is necessary to show that the sequence $\{H_\theta^{(n)}(\ell)\}$ converges to one for each $\theta \in \Omega_A = \Omega_{A,1} \times \Omega_{A,2} \times \ldots \times \Omega_{A,p}$ and each $\ell \in (0, 1)$. Let $\ell \in (0, 1)$ and $\theta \in \Omega_A$, then

$$H_\theta^{(n)}(\ell) = P_\theta[W_n(L_1^{(n)}, \ldots, L_p^{(n)}) \leq w_\ell]$$

where

$$w_\ell = \inf_{-\infty < w < \infty} \{w \mid G_0(w) = \ell\},$$

and $G_0(\cdot)$ is the null distribution function of W_n. Since $G_0(\cdot)$ does not depend on the sample size n, w_ℓ does not depend on n.

Now let $R_\ell = \{(\ell_1, \ldots, \ell_p) \mid W_n(\ell_1, \ldots, \ell_p) \leq w_\ell,\ 0 \leq \ell_i \leq 1\}$

then

$$\ell = \int \cdots \int_{R_\ell} \prod_{i=1}^{p} d\ell_i > 0 ,$$

so there is some $(\ell_1^*, \ell_2^*, \ldots, \ell_p^*) \in R_\ell$ such that $\ell_i^* > 0$, $i = 1, 2, \ldots, p$. Let R^* denote the set

$$R^* = \{(\ell_1, \ldots, \ell_p) \mid \ell_i \leq \ell_i^*,\ i = 1, 2, \ldots, p\} .$$

Because W_n is monotone

$$W_n(\ell_1', \ldots, \ell_p') \leq W_n(\ell_1^*, \ldots, \ell_p^*) \leq w_\ell$$

for $(\ell_1', \ldots, \ell_p') \in R^*$, so

$$R^* \subset R_\ell .$$

Hence, for $\theta \in \Omega_A$,

$$H_\theta^{(n)}(\ell) = P_\theta[R_\ell] \geq P_\theta[R^*]$$

$$= \prod_{i=1}^{p} H_{\theta_i}^{(n)}(\ell_i^*) ,$$

where $H_{\theta_i}^{(n)}(\ell)$ denotes the distribution function of $L_i^{(n)}$ for $\theta_i \in \Omega_{A,i}$. For ε, $0 < \varepsilon < 1$, one can choose N_i, $i = 1, 2, \ldots, p$ such that for $n > N_i$, $H_{\theta_i}^{(n)}(\ell_i^*) > (1 - \varepsilon)^{1/p}$, so for $n > N = \max_{i=1,2,\ldots,p} \{N_i\}$,

$$H_\theta^{(n)}(\ell) > (1 - \varepsilon)$$

or

$$0 \leq 1 - H_\theta^{(n)}(\ell) < \varepsilon$$

and the theorem is proved.

For particular methods of combination it is possible to investigate consistency for the less restrictive combined alternative

$$H_A : \theta_i \in \Omega_{A,i} \quad \text{for at least one} \quad i = 1, 2, \ldots, p . \tag{4.1}$$

For example, let $p = 2$ and suppose the combined alternative is $\theta_1 \in \Omega_{A,1}$, $\theta_2 \in \Omega_{0,2}$. If $\{T_1^{(n)}\}$ is consistent for $H_{0,1}$ versus $H_{A,1}$, then

(i) $L_1^{(n)} \to 0$ in probability $[\theta_1]$, $\theta_1 \in \Omega_{A,1}$,

(ii) $L_2^{(n)}$ is uniform for all n, $\theta_2 \in \Omega_{0,2}$.

Let ε be an arbitrary real number in the interval $(0, 1)$ and consider the region $R = \{(\ell_1, \ell_2) | \ell_1 \ell_2 < \varepsilon\}$ illustrated in Figure 1(a). Now for $\theta = (\theta_1, \theta_2) \in \Omega_A = \Omega_{A,1} \times \Omega_{0,2}$,

$$P_\theta[L_1^{(n)} L_2^{(n)} \leq \varepsilon] \geq P_\theta[L_1^{(n)} \leq \varepsilon] \to 1$$

as $n \to \infty$. Let $L_F^{(n)}$ denote the level of $F_n = L_1^{(n)} L_2^{(n)}$ and $G_0(\cdot)$ the null distribution of F_n; then for each ε, $0 < \varepsilon < 1$,

$$\begin{aligned} P_\theta[L_F^{(n)} \leq \varepsilon] &= P_\theta[G_0(F_n) \leq \varepsilon] \\ &= P_\theta[F_n \leq f_\varepsilon] \to 1 \end{aligned}$$

as $n \to \infty$, where $\varepsilon = G_0(f_\varepsilon)$. Thus $\{F_n\}$ is consistent for H_0 versus $H_A : \theta \in \Omega_{A,1} \times \Omega_{0,2}$.

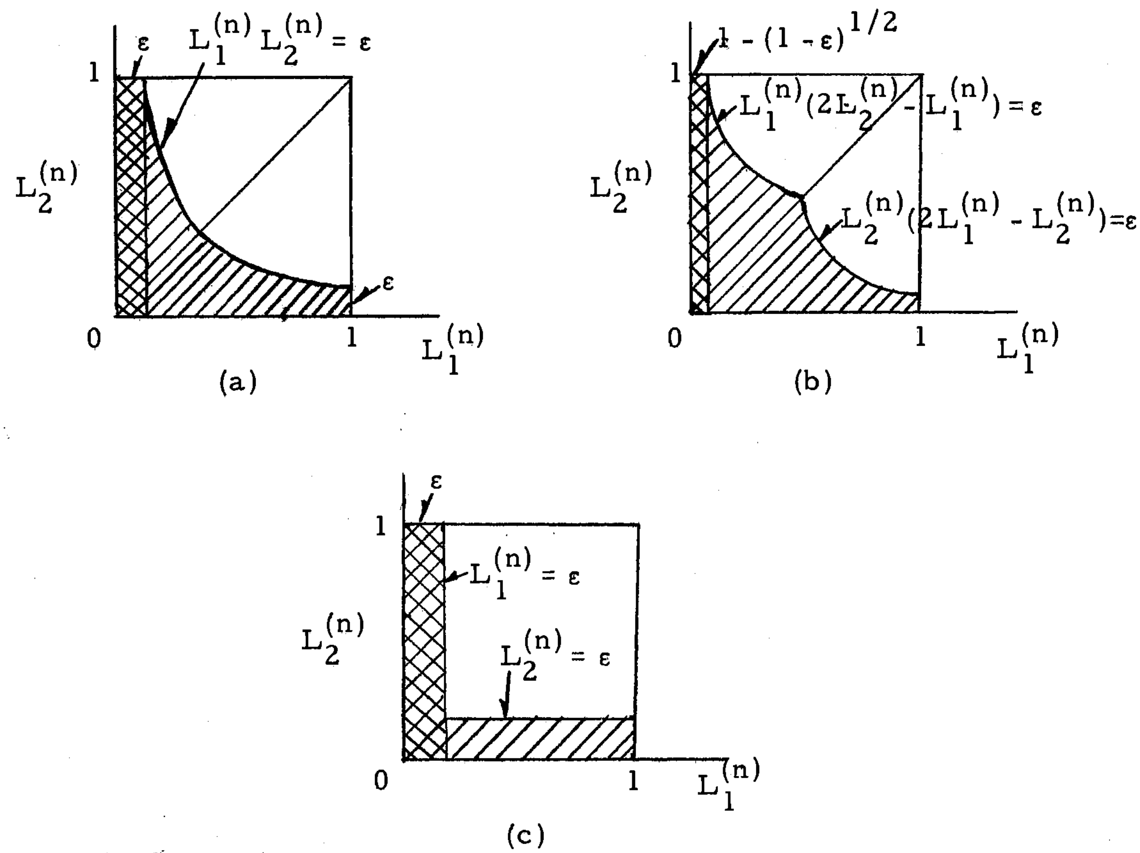

Figure 1. Rejection Regions for Methods of Combination

Similarly, let ε satisfy $0 < \varepsilon < 1$, then from Figure 1(b) we see that

$$P_\theta[L_{(1)}^{(n)}(2L_{(2)}^{(n)} - L_{(1)}^{(n)}) \leq \varepsilon] \geq P_\theta[L_1^{(n)} \leq 1 - (1-\varepsilon)^{1/2}] \to 1, \ \theta \in \Omega_A. \quad (4.2)$$

From Figure 1(c) we see that

$$P_\theta[L_{(1)}^{(n)} \leq \varepsilon] \geq P_\theta[L_1^{(n)} \leq \varepsilon] \to 1, \ \theta \in \Omega_A, \quad (4.3)$$

as $n \to \infty$, where $L_{(1)}^{(n)} = \min(L_1^{(n)}, L_2^{(n)})$ and $L_{(2)}^{(n)} = \max(L_1^{(n)}, L_2^{(n)})$.

Equations (4.2) and (4.3) are sufficient to conclude that for $p = 2$, the joint integral transform method and the method based on the minimum level are consistent for $\theta \in \Omega_A = \Omega_{A,1} \times \Omega_{0,2}$ and $\{T_1^{(n)}\}$ a consistent sequence.

The following theorems will allow us to check the consistency of Pearson's method for H_A (Equation (4.1)) considered above. The first is due to Cramér (9) and the second to Slutsky (43).

<u>Theorem 4.3</u>: Let $\{X_n\}$, $n = 1, 2, 3, \ldots$, be an arbitrary sequence of random variables and let the corresponding sequence of distribution functions $\{F_n(x)\}$ converge to $F(x)$ at every continuity point of $F(x)$. Further let $\{Y_n\}$, $n = 1, 2, 3, \ldots$, be another sequence of random variables which converges in probability to a constant a. Then the sequence of distribution functions of the random variables $X_n Y_n$ converges to the distribution function $F(x/a)$ if $a > 0$ and to the distribution function $1 - F(x/a)$ if $a < 0$.

<u>Theorem 4.4</u>: If the sequences $\{X_1^{(n)}\}, \{X_2^{(n)}\}, \ldots, \{X_r^{(n)}\}, n = 1, 2, \ldots$, of random variables ($r$ is fixed) converge in probability to $a_1, a_2, \ldots, a_r$, respectively, then an arbitrary rational function $R(X_1^{(n)}, \ldots, X_r^{(n)})$, i.e., R is a ratio of polynomials in $X_1^{(n)}, \ldots, X_r^{(n)}$, converges in probability to the constant $R(a_1, a_2, \ldots, a_r)$, provided this constant is finite.

It will be important in the next section to note that,

(i) the sequence $\{X_i^{(n)}\}$ converges to a_i in probability disregarding the other sequences,

(ii) there is no condition of independence for any of the variables involved.

Again let $p = 2$ and consider the alternative $\Omega_A = \Omega_{A,1} \times \Omega_{0,2}$. Then if $\{T_1^{(n)}\}$ is consistent, we can see that Pearson's method

$$P_n = (1 - L_1^{(n)})(1 - L_2^{(n)})$$

converges to a uniform random variable by first applying Theorem 4.4 to $1 - L_1^{(n)}$ and then applying Theorem 4.3 to P_n. Thus Pearson's method is not consistent for $H_A: \theta \in \Omega_{A,1} \times \Omega_{0,2}$. Similar results can be obtained for $p > 2$, but is is felt that the asymptotic property discussed in the following section is of more importance.

Bahadur Slope of Methods of Combination

Littell (28) has compared methods of combination by using the asymptotic theory for comparing test statistics proposed by Bahadur (2). Bahadur's approach is as follows.

Suppose $\{X_n\}$, $n = 1, 2, \ldots$ is a sequence of random variables, each with the same distribution function $F_\theta(\cdot)$, depending on a parameter θ in a set Ω. Let the null hypothesis be

$$H_0: \theta \in \Omega_0 \text{ , } \Omega_0 \subset \Omega \text{ ,}$$

versus the alternative

$$H_A: \theta \in \Omega_A \text{ , } \Omega_A \not\subset \Omega \text{ ,}$$

and

$$\Omega_0 \cap \Omega_A = \emptyset \text{ .}$$

For $n = 1, 2, 3, \ldots$ let $T^{(n)}$ be a real valued statistic for testing H_0 versus H_A which depends only on the first n random variables

$X_1, X_2, \ldots, X_n$. Suppose large values of $T^{(n)}$ are taken to be consistent with the alternative, then the significance level is

$$L^{(n)} = 1 - F_n(T^{(n)})$$

where $F_n(\cdot)$ is the (completely specified) distribution function of $T^{(n)}$ for $\theta \in \Omega_0$. Throughout this section it will be assumed that $T^{(n)}$ is of the continuous type for all $\theta \in \Omega$ and all n.

If the sequence $\{T^{(n)}\}$ is consistent for H_0 versus H_A (see Definition 1.5), then for each $\theta \in \Omega_A$

$$\lim_{n\to\infty} P_\theta[L^{(n)} \leq \ell] = 1 \text{ ,}$$

for all $\ell \in (0, 1)$; that is, the sequence $\{L^{(n)}\}$ converges to zero in probability $[\theta]$, $\theta \in \Omega_A$. Given two consistent sequences of test statistics $\{T_1^{(n)}\}$ and $\{T_2^{(n)}\}$ for the same hypothesis problem, Bahadur's approach is to compute and compare a quantity measuring the rate of convergence of the respective levels to zero. Bahadur (3) has illustrated that it is usually possible, for a sequence of test statistics $\{T^{(n)}\}$, to find a function $c(\theta)$ such that

$$-\frac{2 \log L^{(n)}}{n} \to c(\theta)$$

in probability $[\theta]$, $\theta \in \Omega$, where $c(\theta) = 0$, $\theta \in \Omega_0$, $c(\theta) > 0$, $\theta \in \Omega_A$. The function $c(\theta)$ is called the slope of $\{T^{(n)}\}$ and is used as the measure of the rate of convergence of the sequence $\{L^{(n)}\}$ to zero.

If $\{T_1^{(n)}\}$ and $\{T_2^{(n)}\}$ are two sequences of test statistics for the same hypothesis problem, and if

$$-\frac{2\log L_1^{(n)}}{n} \to c_1(\theta) \quad \text{in probability } [\theta], \ \theta \in \Omega_A ,$$

and

$$-\frac{2\log L_2^{(n)}}{n} \to c_2(\theta) \quad \text{in probability } [\theta], \ \theta \in \Omega_A ,$$

then Bahadur considers $\{T_1^{(n)}\}$ preferable to $\{T_2^{(n)}\}$ if $c_1(\theta) > c_2(\theta)$; i.e., the sequence $\{L_1^{(n)}\}$ converges to zero at a faster rate than does the sequence $\{L_2^{(n)}\}$. For further discussion of $c(\theta)$ with examples and connections between $c(\theta)$ and other asymptotic criteria the reader is referred to Bahadur (2), (3) and Littell (29).

A number of techniques have been devised for the calculation of slopes. For this study the following theorem will be sufficient. For the proof see Savage (40).

<u>Theorem 4.5</u>: If $\{T^{(n)}\}$ is a sequence of test statistics satisfying the properties

(i) there exists a function $b(\theta)$, $0 < b(\theta) < \infty$, such that

$$\frac{T^{(n)}}{\sqrt{n}} \to b(\theta) \quad \text{in probability } [\theta], \ \theta \in \Omega_A ,$$

(ii) there exists a continuous function $f(t)$ such that for each t in some neighborhood of $f(\theta)$,

$$\lim_{n\to\infty} -\frac{1}{n}\log P_0[T^{(n)} \geq \sqrt{n}\, t] = f(t) ,$$

then the slope of $\{T^{(n)}\}$ is given by

$$c(\theta) = 2 f(b(\theta)) .$$

Consider again the combined hypothesis

$$H_0: \theta_i \in \Omega_{0,i}, \quad i = 1, 2, \ldots, p,$$

versus

$$H_A: \theta \in \Omega_{A,i} \text{ for at least one } i = 1, 2, \ldots, p.$$

Now denote the p independent test statistics by $T_1^{(n)}, T_2^{(n)}, \ldots, T_p^{(n)}$ and the respective levels by $L_1^{(n)}, \ldots, L_p^{(n)}$ where n indicates the sample size on which each of the test statistics depend.

Suppose that the sequence $\{T_i^{(n)}\}$ has slope $c_i = c_i(\theta_i)$, $i = 1, 2, \ldots, p$; i.e.,

$$-\frac{2}{n} \log L_i^{(n)} \to c_i > 0 \text{ in probability } [\theta_i], \; \theta_i \in \Omega_{A,i},$$

as $n \to \infty$. The objective is then to compute and compare the slopes of the various methods of combining the p (p remains fixed) independent levels.

Littell obtained the slope for three of the methods of combination mentioned in Chapter I, specifically, for Fisher's method, the minimum of the p levels, and the maximum of the p levels. To illustrate calculation of the slope for a method of combination, consider a slightly modified, yet completely equivalent in terms of the distribution of levels, version of Fisher's statistic defined by

$$F_n = -\frac{2}{\sqrt{n}} \log\{L_1^{(n)} L_2^{(n)} \cdots L_p^{(n)}\}.$$

The slope $c(\theta)$, where $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$, will be calculated using Theorems 4.4 and 4.5. First,

$$\frac{F_n}{\sqrt{n}} = -\frac{2}{n}\log\{L_1^{(n)} L_2^{(n)} \cdots L_p^{(n)}\}$$

$$= -\frac{2}{n}\log L_1^{(n)} - \frac{2}{n}\log L_2^{(n)} - \ldots - 2\log L_p^{(n)}$$

$$\rightarrow c_1(\theta_1) + c_2(\theta_2) + \ldots + c_p(\theta_p) = b(\theta)$$

in probability $[\theta]$, by applying Theorem 4.4 and making use of the assumption

$$-\frac{2\log L_i^{(n)}}{n} \rightarrow c_i(\theta_i) \text{ in probability } [\theta_i] .$$

To calculate $f(t)$ as described in part (ii) of Theorem 4.5, notice that under the null hypothesis F_n is distributed like $Y/\sqrt{n}$ where Y is a chi-squared variable with $2p$ degrees of freedom.

Lemma 4.1: If Y is a chi-squared variable with $2p$ degrees of freedom then

$$\lim_{n\to\infty} -\frac{1}{n} P[Y \geq nt] = \frac{t}{2} , \quad t > 0 .$$

Proof: The variable Y has density function

$$f(y) = \frac{1}{(p-1)!\,2^p} y^{p-1} e^{-y/2} , \quad y > 0$$

$$= 0 , \quad y \leq 0$$

and distribution function

$$F(y) = 0 , \quad y \leq 0$$

$$= 1 - e^{-y/2}\left[\sum_{i=0}^{p-1} y^i/2^i\, i!\right] , \quad y > 0$$

Hence for $t > 0$,

$$\lim_{n\to\infty} -\frac{1}{n}\log P[Y \geq nt] = \lim_{n\to\infty}\left\{-\frac{1}{n}\log\left\{e^{-nt/2}\left[\sum_{i=0}^{p-1}\frac{n^i t^i}{2^i i!}\right]\right\}\right\}$$

$$= \frac{t}{2} - \lim_{n\to\infty}\left\{\frac{\log\left[\sum_{i=0}^{p-1}\frac{n^i t^i}{2^i i!}\right]}{n}\right\}$$

Applying L'Hospital's rule

$$\lim_{n\to\infty} -\frac{1}{n}P[Y \geq nt] = \frac{t}{2} - \lim_{n\to\infty}\left\{\frac{\left[\frac{t}{2} + \frac{t^2 n}{4\cdot 1!} + \frac{t^3 n^2}{8\cdot 2!} + \dots + \frac{t^{p-1}n^{p-2}}{2^{p-1}\cdot(p-2)!}\right]}{\left[1 + \frac{nt}{2\cdot 1!} + \frac{n^2 t^2}{4\cdot 2!} + \dots + \frac{t^{p-1}n^{p-1}}{2^{p-1}(p-1)!}\right]}\right\}$$

$$= \frac{t}{2}\ .$$

Thus $f(t) = \frac{t}{2}$, and Theorem 4.5 gives

$$c(\theta) = 2\cdot f(b(\theta)) = \sum_{i=1}^{p} c_i(\theta_i)\ .$$

That is, the slope of Fisher's method of combination is the sum of the individual slopes.

Similarly, Littell showed that the slope of the combined test based on the maximum level is the number of tests, p, times the minimum of the individual slopes, while the slope of the combined test based on the minimum level is the maximum of the individual slopes. Littell also proved the following theorem:

<u>Theorem 4.6</u>: Suppose $\{T_i^{(n)}\}$ has a maximum slope for testing $H_{0,i}: \theta_i \in \Omega_{0,i}$ versus $H_{A,i}: \theta_i \in \Omega_{A,i}$, $i = 1, 2, \dots, p$. If $c^*(\theta)$ is the

slope of any sequence of tests $\{T_n^*\}$ obtained by combining the levels $\{L_i^{(n)}\}$, $i = 1, 2, \ldots, p$, for testing the combined hypothesis, then

$$c(\theta) \geq c^*(\theta)$$

where $c(\theta)$ is the sum of the individual slopes. In fact, $c(\theta)$ is the maximum slope of all sequences of test statistics for the combined hypothesis problem.

It follows that if each of the sequences $\{T_i^{(n)}\}$ has maximum slope for $H_{0,i}$ versus $H_{A,i}$, then the Fisher method is optimal, among all methods of combining the data, based on the Bahadur slope criteria. The objective here is to compute the slope of Pearson's method and some of the methods constructed in Chapter III.

Several lemmas will first be proved to facilitate calculation of the slopes. In all of the following lemmas

$$0 \leq L_{(1)}^{(n)} \leq L_{(2)}^{(n)} \leq \cdots \leq L_{(p)}^{(n)} \leq 1$$

will denote order statistics arising from $L_1^{(n)}, L_2^{(n)}, \ldots, L_p^{(n)}$, the parameter θ will denote the vector $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$, $\theta_i \in \Omega_i$, $i = 1, 2, \ldots, p$, and $c_{(1)}(\theta) \leq c_{(2)}(\theta) \leq \cdots \leq c_{(p)}(\theta)$ will denote the ordered values of $c_1(\theta_1), c_2(\theta_2), \ldots, c_p(\theta_p)$.

<u>Lemma 4.2</u>: If the random variables $X_1^{(n)}, X_2^{(n)}, \ldots, X_p^{(n)}$ (p fixed) are independent for each $n = 1, 2, 3, \ldots$, and if the sequences $\{X_1^{(n)}\}, \{X_2^{(n)}\}, \ldots, \{X_p^{(n)}\}$ of random variables converge in probability to $a_1, a_2, \ldots, a_p$, respectively, then the sequences $\{X_{(1)}^{(n)}\}, \{X_{(2)}^{(n)}\}, \ldots, \{X_{(p)}^{(n)}\}$ converge to $a_{(1)}, a_{(2)}, \ldots, a_{(p)}$, respectively, where, for each n, $X_{(1)}^{(n)}, X_{(2)}^{(n)}, \ldots, X_{(p)}^{(n)}$ are the order statistics

arising from $X_1^{(n)}, X_2^{(n)}, \ldots, X_p^{(n)}$ and $a_{(1)} \le a_{(2)} \le \cdots \le a_{(p)}$ are the ordered values of $a_1, a_2, \ldots, a_p$.

Proof: For arbitrary $\varepsilon > 0$ and integer k satisfying $1 \le k \le p$,

$$P[|X_{(k)}^{(n)} - a_{(k)}| > \varepsilon] = P[X_{(k)}^{(n)} > a_{(k)} + \varepsilon] + P[X_{(k)}^{(n)} < a_{(k)} - \varepsilon]$$

$$= P[X_{(k)}^{(n)} > b^+] + P[X_{(k)}^{(n)} < b^-]$$

where $b^+ = a_{(k)} + \varepsilon$ and $b^- = a_{(k)} - \varepsilon$. Now

$$P[X_{(k)}^{(n)} > b^+] = P[\text{at least } p\text{-}k\text{+}1 \text{ of } X_1^{(n)}, X_2^{(n)}, \ldots, X_p^{(n)} > b^+]$$

$$= \sum_{i=1}^{k} P[\text{exactly } p\text{-}k\text{+}i \text{ of } X_1^{(n)}, \ldots, X_p^{(n)} > b^+]$$

For an arbitrary integer i satisfying $1 \le i \le k$, the probability

$$P[\text{exactly } p\text{-}k\text{+}i \text{ of } X_1^{(n)}, \ldots, X_p^{(n)} > b^+]$$

can be expressed as the sum of m_i probabilities, where

$$m_i = \frac{p!}{(p-k+i)!\,(k-i)!} \ ,$$

since there are m_i distinct ways of selecting p-k+i random variables from the p variables $X_1, X_2, \ldots, X_p$. With the appropriate renumbering of the p original random variables, any one of these m_i probabilities can be expressed as

$$P[X_1^{(n)} \le b^+, \ldots, X_{k-i}^{(n)} \le b^+,\ X_{k-i+1}^{(n)} > b^+, \ldots, X_p^{(n)} > b^+]$$
$$= \left\{ \prod_{j=1}^{k-i} P[X_j^{(n)} \le b^+] \right\} \left\{ \prod_{j=k-i+1}^{p} P[X_j^{(n)} > b^+] \right\} .$$

The right side of the equation follows because $X_1^{(n)}, X_2^{(n)}, \ldots, X_p^{(n)}$ are independent.

Now since $b^+ = a_{(k)} + \varepsilon$, at most $p-k$ of the $a_1, a_2, \ldots, a_p$ are greater than or equal to b^+, so there is an a_{j^*} among $a_{k-i+1}, a_{k-i+2}, \ldots, a_p$ such that $a_{j^*} < b^+$. Let $\delta = b^+ - a_{j^*} > 0$, then

$$P[X_{j^*}^{(n)} > b^+] = P[X_{j^*}^{(n)} - a_{j^*} > \delta]$$
$$\le P[|X_{j^*}^{(n)} - a_{j^*}| > \delta] \to 0 \quad \text{as} \quad n \to \infty .$$

Since all other probabilities in the product are bounded by zero and one,

$$P[X_1^{(n)} \le b^+, \ldots, X_{k-i}^{(n)} \le b^+, X_{k-i+1}^{(n)} > b^+, \ldots, X_p^{(n)} > b^+] \to 0$$

as $n \to \infty$, which implies

$$P[\text{exactly } p\text{-}k\text{+}i \text{ of } X_1^{(n)}, \ldots, X_p^{(n)} > b^+] \to 0$$

as $n \to \infty$ for all $i = 1, 2, \ldots, k$. Therefore,

$$P[X_{(k)}^{(n)} > b^+] \to 0$$

as $n \to \infty$ for each $k = 1, 2, \ldots, p$. An analogous argument can be used to show that

$$P[X_{(k)}^{(n)} < b^-] \to 0$$

as $n \to \infty$ for each $k = 1, 2, \ldots, p$, and the theorem is established.

Note that if we define $X_i^{(n)} = -\frac{2}{n} \log L_i^{(n)}$, $i = 1, 2, \ldots, p$, then $X_{(p-i+1)}^{(n)} = -\frac{2}{n} \log L_{(i)}^{(n)}$, $i = 1, 2, \ldots, p$, where $X_{(1)}, X_{(2)}, \ldots, X_{(p)}$ denote the order statistics arising from $X_1, X_2, \ldots, X_p$. Thus the following theorem is a direct application of Lemma 4.2.

Theorem 4.7: If $-\frac{2}{n} \log L_i^{(n)} \to c_i(\theta_i)$ in probability $[\theta_i]$, $\theta_i \in \Omega_i$, $i = 1, 2, \ldots, p$, then

$$-\frac{2}{n} \log L_{(i)}^{(n)} \to c_{(p-i+1)}(\theta) \text{ in probability } [\theta] .$$

Lemma 4.3: If the sequence $\{X_n\}$ of random variables is such that $P[0 \leq X_n \leq a] = 1$ for all n, a finite, then

$$-\frac{2}{n} \log (b + X_n) \to 0 \text{ in probability}$$

where $b \geq 1$.

Proof: For $\varepsilon > 0$,

$$\begin{aligned} P[|-\frac{2}{n} \log (b + X_n)| \leq \varepsilon] &= P[-\varepsilon \leq -\frac{2}{n} \log (b + X_n)] \\ &= P[e^{n\varepsilon/2} \geq b + X_n] \\ &= P[X_n \leq e^{n\varepsilon/2} - b] . \end{aligned}$$

Clearly there is an N such that for $n > N$, $e^{n\varepsilon/2} - b > a$, so for $n > N$

$$P[|-\frac{2}{n} \log (b + X_n)| \leq \varepsilon] = 1 .$$

Consider first the joint integral transform method for $p = 2$ independent test statistics $T_1^{(n)}$ and $T_2^{(n)}$. Define the statistic by

$$T_n = -\frac{2}{\sqrt{n}} \log \{L_{(1)}^{(n)} (2 L_{(2)}^{(n)} - L_{(1)}^{(n)})\} .$$

Again there is some modification of the statistic given in Chapter III and again this is to facilitate calculation of the slope. First,

$$\begin{aligned} \frac{T_n}{\sqrt{n}} &= -\frac{2}{n} \log \{L_{(1)}^{(n)} (2 L_{(2)}^{(n)} - L_{(1)}^{(n)})\} \\ &= -\frac{2}{n} \log L_{(1)}^{(n)} - \frac{2}{n} \log L_{(2)}^{(n)} - \frac{2}{n} \log \{1 + 1 - L_{(1)}^{(n)} \big/ L_{(2)}^{(n)}\} . \end{aligned}$$

By Theorem 4, 7,

$$-\frac{2}{n} \log L_{(1)}^{(n)} \to c_{(2)}(\theta) \text{ in probability } [\theta] ,$$

$$-\frac{2}{n} \log L_{(2)}^{(n)} \to c_{(1)}(\theta) \text{ in probability } [\theta] .$$

By Lemma 4.3,

$$-\frac{2}{n} \log \{1 + 1 - L_{(1)}^{(n)} \big/ L_{(2)}^{(n)}\} \to 0 \text{ in probability } [\theta]$$

if we define $X_n = 1 - L_{(1)}^{(n)} \big/ L_{(2)}^{(n)}$, $n = 1, 2, \ldots$. Thus $b(\theta)$ described in Theorem 4.5 is $b(\theta) = c_{(1)}(\theta) + c_{(2)}(\theta)$.

To calculate $f(t)$ as described in part (ii) of Theorem 4.5 we need

$$\begin{aligned} &\lim_{n\to\infty} -\frac{1}{n} \log P_0[T_n \geq \sqrt{n}\, t] \\ &\qquad = \lim_{n\to\infty} -\frac{1}{n} \log P_0 \left[-\frac{2}{\sqrt{n}} \log \left\{ L_{(1)}^{(n)} (2L_{(2)}^{(n)} - L_{(1)}^{(n)}) \right\} \geq \sqrt{n}\, t \right] \\ &\qquad = \lim_{n\to\infty} -\frac{1}{n} \log P_0 \left[L_{(1)}^{(n)} (2L_{(2)}^{(n)} - L_{(1)}^{(n)}) \leq e^{-nt/2} \right] . \end{aligned}$$

By Theorem 3.2,

$$P_0[T_n \geq \sqrt{n}\,t] = 1 - (1 - e^{-nt/2})^{1/2} + \frac{e^{-nt/2}}{2} \log \left[\frac{1 + (1 - e^{-nt/2})^{1/2}}{1 - (1 - e^{-nt/2})^{1/2}} \right].$$

Define $h(n) = P_0[T_n \geq \sqrt{n}\,t]$ and $g(n) = 1 - e^{-nt/2}$, then

$$h(n) = 1 - g^{1/2} + \frac{(1-g)}{2} \log \left[\frac{1 + g^{1/2}}{1 - g^{1/2}} \right].$$

Note that

$$\lim_{n\to\infty} g(n) = 1$$

and

$$\lim_{n\to\infty} h(n) = 0 .$$

Now

$$f(t) = \lim_{n\to\infty} -\frac{1}{n} \log P_0[T_n \geq \sqrt{n}\,t] = \lim_{n\to\infty} -\frac{\log h(n)}{n} = \lim_{n\to\infty} -\frac{h'(n)}{h(n)}$$

After some algebra, $h'(n)$ is found to be

$$h'(n) = -\frac{t}{4}(1 - g) \log \left[\frac{1 + g^{1/2}}{1 - g^{1/2}} \right]$$

so

$$f(t) = \lim_{n\to\infty} -\frac{\log h(n)}{n} = \lim_{n\to\infty} \left\{ \frac{\frac{t}{4}(1 - g) \log \left[\frac{1 + g^{1/2}}{1 - g^{1/2}} \right]}{1 - g^{1/2} + \frac{(1-g)}{2} \log \left[\frac{1 + g^{1/2}}{1 - g^{1/2}} \right]} \right\}$$

$$= \lim_{n\to\infty} \left\{ \frac{\frac{t}{2}}{\frac{2(1 - g^{1/2})}{(1 - g) \log \left[\frac{1 + g^{1/2}}{1 - g^{1/2}} \right]} + 1} \right\}$$

$$= \lim_{n\to\infty} \left\{ \frac{\frac{t}{2}}{\dfrac{2}{(1+g^{1/2})\log\left[\dfrac{1+g^{1/2}}{1-g^{1/2}}\right]} + 1} \right\} = \frac{t}{2} .$$

Theorem 4.5 gives the slope of the joint integral transform method, for $p=2$, to be $2f(b(\theta))$ which is

$$c^*(\theta) = c_{(1)}(\theta) + c_{(2)}(\theta) = c_1(\theta_1) + c_2(\theta_2) ,$$

again optimal based on the Bahadur slope criteria. Without the null distribution function for $p > 2$, it is not possible to compute the slope for general p.

Next consider a statistic equivalent to Pearson's defined by

$$P_n = -\frac{2}{\sqrt{n}} \log\{1 - (1-L_1^{(n)})(1-L_2^{(n)}) \cdots (1-L_p^{(n)})\}$$

In this form large values will be consistent with the alternative. First,

$$\begin{aligned}
\frac{P_n}{\sqrt{n}} &= -\frac{2}{n}\log\{1 - \prod_{i=1}^{p} (1-L_i^{(n)})\} \\
&= -\frac{2}{n}\log\{1 - \prod_{i=1}^{p} (1-L_{(i)}^{(n)})\} \\
&= -\frac{2}{n}\log\{1 - \prod_{i=1}^{p-1} (1-L_{(i)}^{(n)}) + L_{(p)}^{(n)} \prod_{i=1}^{p-1} (1-L_{(i)}^{(n)})\} .
\end{aligned}$$

Define

$$h\left(L_{(1)}^{(n)}, \ldots, L_{(p-1)}^{(n)}\right) = 1 - \prod_{i=1}^{p-1} \left(1 - L_{(i)}^{(n)}\right) ,$$

then

$$\frac{P_n}{\sqrt{n}} = -\frac{2}{n}\log\{h + L_{(p)}^{(n)}(1-h)\}$$

$$= -\frac{2}{n}\log\{L_{(p)}^{(n)} + h - L_{(p)}^{(n)}h\}$$

$$= -\frac{2}{n}\log L_{(p)}^{(n)} - \frac{2}{n}\log\{1 + \frac{h}{L_{(p)}^{(n)}} - h\} .$$

Note that $h\left(L_{(1)}^{(n)}, \ldots, L_{(p-1)}^{(n)}\right)$ satisfies

(i) $P_\theta\left[0 \leq h\left(L_{(1)}^{(n)}, \ldots, L_{(p-1)}^{(n)}\right) \leq 1\right] = 1$

(ii) $h(\ell_{(1)}, \ldots, \ell_{(p-1)}) \geq h(\ell^*_{(1)}, \ldots, \ell^*_{(p-1)})$

for real numbers satisfying $\ell_{(i)} \geq \ell^*_{(i)}$, $i = 1, 2, \ldots, p-1$. If we define the sequence of variables $\{X_n\}$ by

$$X_n = \frac{h\left(L_{(1)}^{(n)}, \ldots, L_{(p-1)}^{(n)}\right)}{L_{(p)}^{(n)}} - h\left(L_{(1)}^{(n)}, \ldots, L_{(p-1)}^{(n)}\right)$$

then for every set of real numbers satisfying

$0 \leq \ell_{(1)} \leq \ell_{(2)} \leq \cdots \leq \ell_{(p)} \leq 1$, x_n satisfies

(i) $x_n = \dfrac{h(\ell_{(1)}, \ldots, \ell_{(p-1)})}{\ell_{(p)}} - h(\ell_{(1)}, \ldots, \ell_{(p-1)}) \geq 0$,

(ii) $x_n = \dfrac{h(\ell_{(1)}, \ldots, \ell_{(p-1)})}{\ell_{(p)}} - h(\ell_{(1)}, \ldots, \ell_{(p-1)})$

$$\leq \frac{h(\ell_{(1)}, \ldots, \ell_{(p-1)})}{\ell_{(p)}} \leq \frac{h(\ell_{(p)}, \ldots, \ell_{(p)})}{\ell_{(p)}}$$

$$= \frac{1-(1-\ell_{(p)})^{p-1}}{\ell_{(p)}} = \frac{[1-(1-\ell_{(p)})]}{\ell_{(p)}}\left[\sum_{i=0}^{p-2}(1-\ell_{(p)})^i\right] \leq p-2$$

Thus $P_\theta[0 \leq X_n \leq p-2] = 1$ for all n, so by Theorem 4.7 and Lemma 4.3,

$$\frac{P_n}{\sqrt{n}} \to c_{(1)}(\theta) \text{ in probability } [\theta].$$

The function $f(t)$ is given by the following Lemma.

<u>Lemma 4.4</u>: If Y is a chi-squared variable with $2p$ degrees of freedom then

$$\lim_{n\to\infty} -\frac{1}{n} \log P[-2\log(1-e^{-Y/2}) \geq nt] = \frac{pt}{2}.$$

Proof: The random variable Y has distribution function

$$F(y) = 1 - e^{-y/2}\left[\sum_{i=0}^{p-1} \frac{y^i}{2^i i!}\right],$$

$$\begin{aligned} P[-2\log(1-e^{-Y/2}) \geq nt] &= P[Y \leq -2\log(1-e^{-nt/2})] \\ &= 1-(1-e^{-nt/2})\left[1+\sum_{i=1}^{p-1} \frac{(-1)^i\{\log(1-e^{-nt/2})\}^i}{i!}\right] \\ &= h(n), \end{aligned}$$

and

$$\begin{aligned} \lim_{n\to\infty} -\frac{1}{n}\log P[Y \leq -2\log(1-e^{-Y/2})] &= \lim_{n\to\infty} -\frac{1}{n}\log h(n) \\ &= \lim_{n\to\infty} -\frac{h'(n)}{h(n)} \\ &= \lim_{n\to\infty} -\frac{h''(n)}{h'(n)} \end{aligned}$$

After some simplification

$$h'(n) = \frac{(-1)^{p-2} t\, e^{-nt/2}}{2(p-1)!} \{\log(1-e^{-nt/2})\}^{p-1} ,$$

$$h''(n) = \frac{(-1)^{p-2} t^2 e^{-nt/2}}{4(p-1)!} \{\log(1-e^{-nt/2})\}^{p-2} \left\{ \frac{(p-1)e^{-nt/2}}{(1-e^{-nt/2})} - \log(1-e^{-nt/2}) \right\}$$

and

$$\lim_{n\to\infty} - \frac{h''(n)}{h'(n)} = \lim_{n\to\infty} - \frac{t}{2} \left\{ \frac{(p-1)e^{-nt/2}}{(1-e^{-nt/2})\log(1-e^{-nt/2})} - 1 \right\}$$

$$= - \frac{(p-1)t}{2} \left\{ \frac{1}{\lim\limits_{n\to\infty} e^{nt/2} \log(1-e^{-nt/2})} \right\} + \frac{t}{2} .$$

Now

$$\lim_{n\to\infty} e^{nt/2} \log(1-e^{-nt/2}) = \lim_{n\to\infty} \frac{\log(1-e^{-nt/2})}{e^{-nt/2}}$$

$$= \lim_{n\to\infty} \frac{\frac{t}{2} e^{-nt/2}}{-\frac{t}{2} e^{-nt/2}(1-e^{-nt/2})} = -1 .$$

Hence

$$\lim_{n\to\infty} - \frac{1}{n} \log P[-2\log(1-e^{-Y/2}) \geq nt] = \frac{pt}{2}$$

and the lemma is proved.

Applying this lemma

$$\lim_{n\to\infty} - \frac{1}{n} P_0[P_n \geq \sqrt{n}\, t] = \lim_{n\to\infty} - \frac{1}{n} P_0[-2\log(1-e^{-Y/2}) \geq nt] = \frac{pt}{2}$$

since under the combined null hypothesis, $-2\log\{\prod_{i=1}^{p}(1-L_{(i)}^{(n)})\}$ is a chi-squared variable with 2p degrees of freedom. Finally, the slope of Pearson's method is

$$c^{**}(\theta) = 2\left(\frac{p\,c_{(1)}}{2}\right) = p\,c_{(1)}(\theta) ,$$

or the number of levels, p, times the minimum of the individual slopes.

Statistics suggested in the last section of Chapter III can also be used as methods of combining independent levels. However, general expressions for the slope can be obtained only if some restrictions are placed on the combined alternative and individual slopes. First, it will be assumed that the combined alternative is

$$H_A^*: \theta_i \in \Omega_{A,i} \quad \text{for all} \quad i = 1, 2, \ldots, p ,$$

specifically because it will be necessary to have $c_{(i)}(\theta) > 0$ for $\theta = (\theta_1, \ldots, \theta_p) \in \Omega_A^*$, where now

$$\Omega_A^* = \Omega_{A,1} \times \Omega_{A,2} \times \cdots \times \Omega_{A,p} .$$

From a practical point of view one might hypothesize a value of a parameter near the actual value and hopefully this will lead to small value for the slope; however,it is unlikely that one will hypothesize the exact value of any parameter. Thus it is felt this assumption does not greatly restrict the results.

The second assumption is that $c_{(i)}(\theta) \neq c_{(j)}(\theta)$, $i \neq j$, for the θ of interest in $\theta \in \Omega_A$. This assumption will place some restriction on the application of the results. However, in the important case when the $T_i^{(n)}$, $i = 1, 2, \ldots, p$, are distinct test statistics of the same or distinct hypotheses the assumption is very reasonable.

With these assumptions three more lemmas hold.

Lemma 4.5: For $\theta \in \Omega_A^*$, $L_{(i)}^{(n)} \to 0$ in probability $[\theta]$.

Proof: If we select an $\varepsilon > 1$, then $P_\theta[L_{(i)}^{(n)} < \varepsilon] = 1$. If we select an ε satisfying $0 < \varepsilon < 1$ then

$$P_\theta\left[L_{(i)}^{(n)} \leq \varepsilon\right] = P_\theta\left[-\frac{2\log L_{(i)}^{(n)}}{n} > -\frac{2\log\varepsilon}{n}\right]$$

$$= P_\theta\left[-\frac{2\log L_{(i)}^{(n)}}{n} - c_{(p-i+1)} \geq -c_{(p-i+1)} - \frac{2\log\varepsilon}{n}\right].$$

Since $0 < \varepsilon < 1$, $-\dfrac{2\log\varepsilon}{n} > 0$, but since $c_{(p-i+1)} > 0$ we can find an N such that for $n \geq N$, $-c_{(p-i+1)} - \dfrac{2\log\varepsilon}{n} < 0$. So for $n \geq N$,

$$P_\theta\left[L_{(i)}^{(n)} \leq \varepsilon\right] = P_\theta\left[-\frac{2\log L_{(i)}^{(n)}}{n} - c_{(p-i+1)} \geq -c_{(p-i+1)} - \frac{2\log\varepsilon}{n}\right]$$

$$\geq P_\theta\left[\left|-\frac{2\log L_{(i)}^{(n)}}{n} - c_{(p-i+1)}\right| \leq c_{(p-i+1)} + \frac{2\log\varepsilon}{n}\right]$$

$$\geq P_\theta\left[\left|-\frac{2\log L_{(i)}^{(n)}}{n} - c_{(p-i+1)}\right| \leq c_{(p-i+1)} + \frac{2\log\varepsilon}{N}\right]$$

$$\to 1 \text{ as } n \to \infty.$$

Lemma 4.6: For $\theta \in \Omega_A^*$,

$$\frac{L_{(i)}^{(n)}}{L_{(j)}^{(n)}} \to 0 \text{ in probability } [\theta]$$

for each $i = 1, 2, \ldots, p-1; j = 2, \ldots, p; j > i$.

Proof: Since $c_{(p-i+1)} > c_{(p-j+1)}$,

$$-\frac{2}{n}\log\left[\frac{L_{(i)}^{(n)}}{L_{(j)}^{(n)}}\right] = -\frac{2}{n}\log L_{(i)}^{(n)} + \frac{2}{n}\log L_{(j)}^{(n)}$$

$$\to c_{(p-i+1)} - c_{(p-j+1)} > 0 \text{ in probability } [\theta],\ \theta \in \Omega_A^* .$$

So for $0 < \varepsilon < 1$, $k = c_{(p-i+1)} - c_{(p-j+1)}$, and $n > -\frac{4\log\varepsilon}{k}$, it follows that

$$P_\theta\left\{\frac{L_{(i)}^{(n)}}{L_{(j)}^{(n)}} < \varepsilon\right\} \geq P_\theta\left\{\frac{L_{(i)}^{(n)}}{L_{(j)}^{(n)}} < e^{-(\frac{n}{2})(\frac{k}{2})}\right\}$$

$$= P_\theta\left\{-\frac{2}{n}\log\frac{L_{(i)}^{(n)}}{L_{(j)}^{(n)}} > \frac{k}{2}\right\}$$

$$= P_\theta\left\{-\frac{2}{n}\log\frac{L_{(i)}^{(n)}}{L_{(j)}^{(n)}} - k > -\frac{k}{2}\right\}$$

$$\geq P_\theta\left\{\left|-\frac{2}{n}\log\frac{L_{(i)}^{(n)}}{L_{(j)}^{(n)}} - k\right| < \frac{k}{2}\right\}$$

$$\to 1 \text{ as } n \to \infty .$$

Lemma 4.7: If the sequence of random variables $\{X_n\}$ converges to zero in probability then the sequence

$$\{-\frac{2}{n}\log(a + X_n)\}$$

converges to zero in probability, where $a > 0$ is constant.

Proof: For $\varepsilon > 0$,

$$P\{|-\frac{2}{n}\log(a+X_n)| \leq \varepsilon\} = P\{-\varepsilon \leq -\frac{2}{n}\log(a+X_n) \leq \varepsilon\}$$

$$= P\{e^{-n\varepsilon/2} \leq a+X_n \leq e^{n\varepsilon/2}\}$$

$$= P\{-a+e^{-n\varepsilon/2} \leq X_n \leq -a+e^{n\varepsilon/2}\}$$

Clearly there is some N and some $b > 0$ such that for $n > N$, $-a+e^{-n\varepsilon/2} < -b$ and $-a+e^{n\varepsilon/2} > b$, so for $n > N$

$$P\{|-\frac{2}{n}\log(a+X_n)| \leq \varepsilon\} \geq P\{|X_n| \leq b\}$$

$$\rightarrow 1 \text{ as } n \rightarrow \infty .$$

Now consider the statistic V_n, equivalent to W_4 defined in Chapter III, given by

$$\frac{V_n}{\sqrt{n}} = -\frac{2}{n}\log\left[\left\{1-(1-L_{(1)})^p\right\}\prod_{i=1}^{p-1}\left\{1-\frac{(1-L_{(i+1)})^{p-i}}{(1-L_{(i)})^{p-i}}\right\}\right]$$

$$= -\frac{2}{n}\sum_{i=1}^{p-1}\log\{(1-L_{(i)})^{p-i}-(1-L_{(i+1)})^{p-i}\}$$

$$+\frac{2}{n}\sum_{i=1}^{p-1}\log\{(1-L_{(i)})^{p-i}\}$$

$$-\frac{2}{n}\log\{1-(1-L_{(1)})^p\} .$$

The superscript (n) is deleted here to avoid confusion with powers.

For $i = 1, 2, \ldots, p-2$, $p > 2$,

$$(1 - L_{(i)})^{p-i} - (1 - L_{(i+1)})^{p-i}$$

$$= 1 + \sum_{k=1}^{p-i} \binom{p-i}{k} (-1)^k L_{(i)}^k - 1 - \sum_{k=1}^{p-i} \binom{p-i}{k} (-1)^k L_{(i+1)}^k$$

$$= (p-i) L_{(i+1)} - \sum_{k=2}^{p-i} \binom{p-i}{k} (-1)^k L_{(i+1)}^k + \sum_{k=1}^{p-i} \binom{p-i}{k} (-1)^k L_{(i)}^k$$

$$= L_{(i+1)} \left\{ (p-i) - \sum_{k=2}^{p-i} \binom{p-i}{k} (-1)^k L_{(i+1)}^{k-1} + \frac{L_{(i)}}{L_{(i+1)}} \sum_{k=1}^{p-i} \binom{p-i}{k} (-1)^k L_{(i)}^{k-1} \right\} \tag{4.4}$$

Since $L_{(i)} \to 0$ in probability $[\theta]$ by Lemma 4.5, $L_{(i)}/L_{(i+1)} \to 0$ in probability $[\theta]$ by Lemma 4.6, then by applying Lemma 4.7 to $-\frac{2}{n}\log$ of Equation (4.4).

$$-\frac{2}{n} \log \{(1 - L_{(i)})^{p-i} - (1 - L_{(i+1)})^{p-i}\} \to c_{(p-i)} \text{ in probability } [\theta],$$
$$i = 1, 2, \ldots, p-2 .$$

For $i = p-1$,

$$-\frac{2}{n} \log \{(1 - L_{(p-1)}) - (1 - L_{(p)})\} = -\frac{2}{n} \log \{L_{(p)}\} - \frac{2}{n} \log \{1 - \frac{L_{(p-1)}}{L_{(p)}}\}$$
$$\to c_{(1)}(\theta) \text{ in probability } [\theta] .$$

Now

$$-\frac{2}{n} \log \{1 - (1 - L_{(1)})^p\} = -\frac{2}{n} \log \{1 - 1 - \sum_{k=1}^{p} \binom{p}{k} (-1)^k L_{(1)}^k\}$$

$$= -\frac{2}{n} \log L_{(1)} - \frac{2}{n} \log \{1 - \sum_{k=2}^{p} \binom{p}{k} (-1)^k L_{(1)}^{k-1}\}$$

$$\to c_{(p)} \text{ in probability } [\theta]$$

by again applying Lemmas 4.5 and 4.7. The sequence

$$\frac{2}{n} \sum_{i=1}^{p-1} \log\{(1 - L_{(i)})^{p-i}\} = \frac{2(p-i)}{n} \sum_{i=1}^{p-i} \log(1 - L_{(i)})$$

$$\rightarrow 0 \text{ in probability } [\theta]$$

by Lemma 4.7. Hence

$$\frac{V_n}{\sqrt{n}} \rightarrow \sum_{i=1}^{p} c_{(i)}(\theta) \text{ in probability } [\theta].$$

The function $f(t)$ is provided by Lemma 4.1 because under the combined null hypothesis, $\sqrt{n}\, V_n$ is a chi-squared variable with $2p$ degrees of freedom; i.e., $f(t) = t/2$.

Finally the slope of V_n, say $c^{***}(\theta)$, is

$$c^{***}(\theta) = \frac{2\left\{\sum_{i=1}^{p} c_{(i)}(\theta)\right\}}{2} = \sum_{i=1}^{p} c_i(\theta_i), \quad \theta \in \Omega_A^*.$$

It is important to note that V_n is not a monotone method of combination, yet does possess optimal Bahadur slope.

A summary of the results of this section is included in Chapter VI. It is clear, however, that the joint integral transform approach does yield methods of combination which share the optimal Bahadur slope with Fisher's method for many important problems.

CHAPTER V

GOODNESS OF FIT

Techniques similar to those used to construct methods of combination in Chapter III will now be applied to the general goodness of fit hypothesis problem. To restate the problem, suppose that for a continuous random variable X with distribution function $F(\cdot)$ it is of interest to test the null hypothesis

$$H_0: F(x) = F_0(x), \quad x \in (-\infty, \infty), \tag{5.1}$$

versus the alternative

$$H_A: F(x) \neq F_0(x) \tag{5.2}$$

for some $x \in (-\infty, \infty)$. Throughout this chapter the following assumptions and notation will be used:

(i) $X_1, X_2, \ldots, X_n$ will denote n mutually independent identically distributed continuous random variables.

(ii) $U_1, U_2, \ldots, U_n$ will denote the integral transforms of $X_1, X_2, \ldots, X_n$, respectively; i.e., $U_k = F_0(X_k)$, $k = 1, 2, \ldots, n$.

(iii) $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$ will denote the order statistics arising from $U_1, U_2, \ldots, U_n$.

A Combining Levels Approach to

Goodness of Fit

Suppose we consider each random variable X_k as a test statistic for the hypothesis

$$H_{0,k}: F_k(x) = F_0(x), \text{ all } x \in (-\infty, \infty)$$

versus

$$H_{A,k}: F_k(x) \neq F_0(x), \text{ for some } x \in (-\infty, \infty)$$

where $F_k(\cdot)$ denotes the distribution function of X_k. Then the goodness of fit hypothesis, Equations (5. 1) and (5. 2), is the combined hypothesis

$$H_0: F_k(x) = F_0(x), \ x \in (-\infty, \infty), \ k = 1, 2, \ldots, n,$$

versus the combined alternative

$$H_A: F_k(x) \neq F_0(x), \text{ for some } x \in (-\infty, \infty), \ k = 1, 2, \ldots, n.$$

If it were possible to define an appropriate significance level, say L_k, for each X_k, then perhaps we could use Fisher's method

$$Q_1 = -2 \sum_{k=1}^{n} \log L_k$$

to test H_0 versus H_A. By an "appropriate" level we mean that small values should reflect departure from the null hypothesis in the direction of the alternative.

In Chapters II and III it was possible to define appropriate levels because of the a priori knowledge that small values of each test statistic are consistent with its respective alternative. Such a priori knowledge is not available for the goodness of fit hypothesis. One might suggest

$$L_k = 2\,\mathrm{Min}\,\{F_0(X_k),\ 1 - F_0(X_k)\}\ ,$$

the definition generally used for two-tailed test statistics when it is known that either large or small values of the statistic are consistent with the alternative. Note that L_k as defined here is a uniform variable under the null hypothesis $F(x) = F_0(x)$, $x \in (-\infty, \infty)$. This definition seems inappropriate for H_0 versus H_A, but it does suggest a possible approach.

Consider the order statistics $U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n)}$ arising from $U_k = F_0(X_k)$, $k = 1, 2, \ldots, n$. Under the null hypothesis the $U_{(k)}$ are uniform order statistics. Under any alternative it seems reasonable to assume that, for at least some of these variables, either smaller or larger values than predicted by uniform are likely. Perhaps a reasonably sensitive statistic would be

$$Q_2 = -2 \sum_{k=1}^{n} \log L_k^*$$

where $L_k^* = 2\,\mathrm{Min}\,\{G_0^{(k)}(U_{(k)}),\ 1 - G_0^{(k)}(U_{(k)})\}$ and $G_0^{(k)}(\cdot)$ denotes the null distribution function of $U_{(k)}$. The function $2\,\mathrm{Min}\,\{G_0^{(k)}(U_{(k)}),\ 1 - G_0^{(k)}(U_{(k)})\}$ would attain its maximum value if $U_{(k)}$ were to realize its median value (under H_0), and the function decreases as $U_{(k)}$ either decreases or increases from its median value. Thus small values of L_k^*, large value of Q_2, should be taken consistent with the alternative.

Each of the L_k^* is a uniform variable under H_0, but the L_k^* are dependent random variables and therefore Q_2 is not a chi-squared variable under H_0. Rather than attempt to find the null distribution of Q_2, suppose we turn to the conditional approach used in the last section of Chapter III. Define the conditional level, say $L_{k|k+1}$, of $U_{(k)}$ to be

$$L_{k|k+1} = 2\,\mathrm{Min}\,\{G_0^{(k|k+1)}(U_{(k)}|U_{(k+1)}),\ 1 - G_0^{(k|k+1)}(U_{(k)})|U_{(k+1)})\}\ ,\qquad k = 1, 2, \ldots, n-1\ ,$$

$$L_{n|n+1} = 2\,\mathrm{Min}\,\{G_0^{(n)}(U_{(n)}),\ 1 - G_0^{(n)}(U_{(n)})\}\ ,$$

where $G_0^{(k|k+1)}(\cdot\,|\cdot)$ is the conditional distribution function (under H_0) of $U_{(k)}$ given $U_{(k+1)}$. Recalling previous discussion of uniform order statistics this is actually the conditional distribution function of $U_{(k)}$ given $U_{(k+1)}, U_{(k+2)}, \ldots, U_{(n)}$. Hence by Theorem 3, 5 $L_{k|k+1}$, $k = 1, 2, \ldots, n$ are mutually independent uniform variables under H_0. The conditional distribution function is

$$G_0^{(k|k+1)}(U_{(k)}|U_{(k+1)}) = (U_{(k)}/U_{(k+1)})^k,\ 0 \le U_{(k)} \le U_{(k+1)}\ ,\qquad k = 1, 2, \ldots, n-1$$

$$G_0^{(n)}(U_{(n)}) = (U_{(n)})^n,\ 0 \le U_{(n)} \le 1\ ,$$

and

$$L_{k|k+1} = 2\,\mathrm{Min}\,\{(U_{(k)}/U_{(k+1)})^k,\ 1 - (U_{(k)}/U_{(k+1)})^k\}\ ,\qquad k = 1, 2, \ldots, n$$

if we define $U_{(n+1)} = 1$. The test statistic Q_3 defined by

$$Q_3 = -2 \sum_{k=1}^{n} \log L_{k|k+1}$$

is a chi-squared variable with $2n$ degrees of freedom under H_0.

Small values of $L_{k|k+1}$, large values of Q_3, will again be taken consistent with the alternative. Because of the conditional nature of $L_{k|k+1}$, this approach has perhaps lost some of its appeal. In fact it can be demonstrated that Q_3 is a biased test statistic for at least one alternative. Consider the class of alternatives.

$$H_A : F(x) = [F_0(x)]^{\alpha}, \ \alpha > 0 .$$

Then for $U_k = F_0(X_k)$ and $0 \le u_k \le 1$

$$\begin{aligned} P_A\{U_k \le u_k\} &= P_A\{F_0(X_k) \le u_k\} \\ &= P_A\{[F_0(X_k)]^{\alpha} \le u_k^{\alpha}\} \\ &= u_k^{\alpha}, \end{aligned}$$

and the joint density of $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$ is

$$\begin{aligned} f_A(u_{(1)}, u_{(2)}, \ldots, u_{(n)}) &= \alpha^n n! \prod_{k=1}^{n} u_{(k)}^{\alpha-1}, \ 0 \le u_{(1)} \le \ldots \le u_{(n)} \le 1 \\ &= 0, \text{ otherwise} \end{aligned}$$

If we define $U_{(n+1)} = 1$ and $Y_k = (U_{(k)} / U_{(k+1)})^k$, $k = 1, 2, \ldots, n$, then the joint density of $Y_1, Y_2, \ldots, Y_n$ is

$$\begin{aligned} f_A(y_1, y_2, \ldots, y_n) &= \alpha^n \prod_{k=1}^{n} y_k^{\alpha-1}, \ 0 \le y_k \le 1, \ k = 1, 2, \ldots, n, \\ &= 0, \text{ otherwise}, \end{aligned}$$

Note that the Y_k are independent, and in terms of $Y_1, \ldots, Y_n$, $L_{k|k+1}$ is

$$L_{k|k+1} = 2\,\mathrm{Min}\{Y_k,\ 1 - Y_k\},\quad k = 1, 2, \ldots, n\ .$$

Hence for $0 \le \ell \le 1$

$$\begin{aligned} P_A[L_{k|k+1} \le \ell] &= P_A[2\,\mathrm{Min}\{Y_k,\ 1 - Y_k\} \le \ell] \\ &= P_A[Y_k \le \tfrac{\ell}{2}] + P_A[Y_k \ge 1 - \tfrac{\ell}{2}] \\ &= (\tfrac{\ell}{2})^\alpha + 1 - (1 - \tfrac{\ell}{2})^\alpha\ . \end{aligned}$$

The null hypothesis corresponds to $\alpha = 1$, and in this case

$$P_0[L_{k|k+1} \le \ell] = \ell\ .$$

However, if $\alpha = 2$ the distribution of X differs significantly for the null hypothesis, but

$$\begin{aligned} P_A[L_{k|k+1} \le \ell] &= \frac{\ell^2}{4} + 1 - 1 + \ell - \frac{\ell^2}{4} \\ &= \ell\ , \end{aligned}$$

and

$$H_A^{(3)}(\ell_3) = H_0^{(3)}(\ell_3) = \ell_3$$

where $H^{(3)}(\cdot)$ denotes the distribution function of the significance level of Q_3. Hence Q_3 is only trivially unbiased for the alternative $\alpha = 2$.

Some 1000 samples of size $n = 15$ from the alternative

$$H_A : F(x) = [F_0(x)]^{3/2}$$

where synthetically generated, and for each, Q_3 was calculated. Since the null distribution of Q_3 is chi-squared with 30 degrees of freedom,

$$0.10 = P_0[Q_3 \geq 40.26]$$

(see any numerical tabulation of the chi-squared distribution) and

$$H_A^{(3)}(.1) = P_A[Q_3 \geq 40.26].$$

Of the 1000 values of Q_3, 68 or 6.8 percent of the values exceeded 40.26. Considering this as an estimate of a proportion p based on 1000 samples, the probability is approximately .95 that p is in the interval

$$.068 - 1.96\sqrt{\frac{(.068)(.932)}{1000}} = .053$$

to

$$.068 + 1.96\sqrt{\frac{(.068)(.932)}{1000}} = .083$$

Hence, this is a good indication that

$$H_A^{(3)}(.1) < .1$$

and that Q_3 is a biased test for this alternative. Note that for $\alpha = 3/2$,

$$E_A[U_k] = 3/2\int_0^1 u^{3/2}\,du = 0.6 > E_0[U_k] = 0.5.$$

Further empirical investigation with a variety of alternatives indicate that Q_3 generally provides poor sensitivity when the mean of the alternative is larger than that of the null. Some of the results are given in the last section of this chapter. In this same section it will also be demonstrated that Q_3 provides extremely good sensitivity for an important and practical problem.

Recalling the discussion given in the last section of Chapter III, many other definitions of conditional significance levels of the $U_{(k)}$ are possible. For example, consider the random variables, $Z_1, Z_2, \ldots, Z_n$ where observed values are defined by

$$z_1 = P_0[U_{(1)} \geq u_{(1)}] = (1 - u_{(1)})^n$$

$$z_2 = P_0[U_{(2)} \geq u_{(2)} | U_{(1)} = u_{(1)}] = \frac{(1 - u_{(2)})^{n-1}}{(1 - u_{(1)})^{n-1}}$$

$$z_k = P_0[U_{(k)} \geq u_{(k)} | U_{(i)} = u_{(i)},\ i = 1, 2, \ldots, k-1]$$

$$= \frac{(1 - u_{(k)})^{n-k+1}}{(1 - u_{(k-1)})^{n-k+1}}, \quad k = 2, 3, \ldots, n .$$

Now if we define the conditional levels to be

$$L_{1|0} = 2\,\mathrm{Min}\{Z_1, 1 - Z_1\},$$

$$L_{k|k-1} = 2\,\mathrm{Min}\{Z_k, 1 - Z_k\}, \quad k = 2, 3, \ldots, n,$$

then again the $L_{k|k-1}$, $k = 1, 2, \ldots, n$ are mutually independent uniform variables under H_0. Empirical studies of

$$Q_4 = -2 \sum_{k=1}^{n} \log L_{k|k-1}$$

indicate that this statistic is extremely sensitive to alternatives with larger mean than the null, while less sensitive to alternatives with smaller mean than the null.

Again referring to Chapter III, it is not necessary to begin the process of defining conditional levels with either $U_{(1)}$ or $U_{(n)}$. The variables $Z_1, Z_2, \ldots, Z_n$ will be independent uniform variables (under H_0) if we observe the following rules:

(i) $z_1 = P_0\{U_{(k)} = u_{(k)} | U_{(j)} = u_{(j)},\ j = 1, 2, \ldots, n; j \neq k\}$ for any $k = 1, 2, \ldots, n$;

(ii) Once we define a z as the conditional (null) probability that $U_{(k)}$ is less than or equal to $u_{(k)}$, $U_{(k)}$ does not appear again in the process ;

(iii) Until we define a z as the conditional (null) probability that $U_{(k)}$ is less than or equal to $u_{(k)}$, all probabilities will be conditioned on $U_{(k)} = u_{(k)}$.

For each sample size n, $n!$ different definitions of the Z's are possible, two of which lead to Q_3 and Q_4 defined above. One more will now be considered here. Let us begin with $n = 4$. For observed $u_{(1)}$, $u_{(2)}$, $u_{(3)}$ and $u_{(4)}$ define z_1, z_2, z_3, and z_4 by

$$z_1 = P_0\{U_{(2)} \geq u_{(2)} | U_{(1)} = u_{(1)},\ U_{(3)} = u_{(3)},\ U_{(4)} = u_{(4)}\}$$

$$= \frac{1}{u_{(3)} - u_{(1)}} \int_{u_{(2)}}^{u_{(3)}} dt = \frac{u_{(3)} - u_{(2)}}{u_{(3)} - u_{(1)}} ,$$

$$z_2 = P_0\{U_{(3)} \leq u_{(3)} | U_{(1)} = u_{(1)},\ U_{(4)} = u_{(4)}\}$$

$$= \frac{2}{(u_{(4)} - u_{(1)})^2} \int_{u_{(1)}}^{u_{(3)}} (t - u_{(1)})\, dt = \frac{(u_{(3)} - u_{(1)})^2}{(u_{(4)} - u_{(1)})^2} \ ,$$

$$z_3 = P_0\{U_{(1)} \geq u_{(1)} \mid U_{(4)} = u_{(4)}\}$$

$$= \frac{3}{u_{(4)}^3} \int_{u_{(1)}}^{u_{(4)}} (u_{(4)} - t)^2\, dt = \frac{(u_{(4)} - u_{(1)})^3}{u_{(4)}^3} \ ,$$

$$z_4 = P_0\{U_{(4)} \leq u_{(4)}\} = \int_0^{u_{(4)}} 4t^3\, dt = u_{(4)}^4 \ .$$

Now define

$$L_1^{**} = 2\,\mathrm{Min}\,\{Z_1 , 1 - Z_1\} \ ,$$

$$L_2^{**} = 2\,\mathrm{Min}\,\{Z_2 , 1 - Z_2\} \ ,$$

$$L_3^{**} = 2\,\mathrm{Min}\,\{Z_3 , 1 - Z_3\} \ ,$$

$$L_4^{**} = 2\,\mathrm{Min}\,\{Z_4 , 1 - Z_4\} \ .$$

Actually L_1^{**} is designed to measure departure of $U_{(2)}$ from that predicted by H_0, L_2^{**} is designed to measure departure of $U_{(3)}$, etc., but it will be convenient to use the difference in ranks of the u's in the numerator as the subscript for the corresponding significance level. Also note that in the definitions of z_1 and z_3 the inequalities are opposite those used to define z_2 and z_4. This does not affect the null distribution of the Z's, and the L's are exactly the same variables as if the inequalities were in the other direction. The purpose is to help establish a recognizable pattern in the definitions.

For $n = 5$, define

$$z_1 = P_0\{U_{(3)} \leq u_{(3)} | U_{(2)} = u_{(2)},\ U_{(4)} = u_{(4)}\} = \frac{u_{(3)} - u_{(2)}}{u_{(4)} - u_{(2)}} ,$$

$$z_2 = P_0\{U_{(2)} \geq u_{(2)} | U_{(1)} = u_{(1)},\ U_{(4)} = u_{(4)}\} = \frac{(u_{(4)} - u_{(2)})^2}{(u_{(4)} - u_{(1)})^2} ,$$

$$z_3 = P_0\{U_{(4)} \leq u_{(4)} | U_{(1)} = u_{(1)},\ U_{(5)} = u_{(5)}\} = \frac{(u_{(4)} - u_{(1)})^3}{(u_{(5)} - u_{(1)})^3} ,$$

$$z_4 = P_0\{U_{(1)} \geq u_{(1)} | U_{(5)} = u_{(5)}\} = \frac{(u_{(5)} - u_{(1)})^4}{u_{(5)}^4}$$

$$z_5 = P_0\{U_{(5)} \leq u_{(5)}\} = u_{(5)}^5 .$$

Now define $L_k^{**} = 2\,\mathrm{Min}\{Z_k,\ 1 - Z_k\}$, $k = 1, 2, \ldots, 5$.

Before considering the general case, let us define $U_{(k)} = 0$. $k < 1$ and $U_{(k)} = 1$, $k > n$. Now for observed $u_{(k)}$, $k = 1, 2, \ldots, n$, $0 \leq u_{(1)} \leq \ldots \leq u_{(n)} \leq 1$, and for any $k = 1, 2, \ldots, n$ and $r > k$,

$$P_0\{U_{(k)} \geq u_{(k)} | U_{(1)} = u_{(1)}, \ldots, U_{(k-1)} = u_{(k-1)},\ U_{(r)} = U_{(r)}, \ldots, U_{(n)} = u_{(n)}\}$$

$$= P_0\{U_{(k)} \geq u_{(k)} | U_{(k-1)} = u_{(k-1)},\ U_{(r)} = u_{(r)}\}$$

$$= \frac{(r-k)}{(u_{(r)} - u_{(k-1)})^{r-k}} \int_{u_{(k)}}^{u_{(r)}} (u_{(r)} - t)^{r-k-1}\, dt$$

$$= \frac{(u_{(r)} - u_{(k)})^{r-k}}{(u_{(r)} - u_{(k-1)})^{r-k}} ;$$

for any $k = 1, 2, \ldots, n$ and $r < k$

$$P_0\{U_{(k)} \leq u_{(k)} | U_{(1)} = u_{(1)}, \ldots, U_{(r)} = u_{(r)}, U_{(k+1)} = u_{(k+1)}, \ldots, U_{(n)} = u_{(n)}\}$$

$$= P_0\{U_{(k)} \leq u_{(k)} | U_{(r)} = u_{(r)}, U_{(k+1)} = U_{(k+1)}\}$$

$$= \frac{(k-r)}{(u_{(k+1)} - u_{(r)})^{k-r}} \int_{u_{(r)}}^{u_{(k)}} (t - r_{(r)})^{k-r-1} dt$$

$$= \frac{(u_{(k)} - u_{(r)})^{k-r}}{(u_{(k+1)} - u_{(r)})^{k-r}} \quad .$$

For general n, n even, we define

$$z_1 = P_0\{U_{(n/2)} \geq u_{(n/2)} | \text{all other } U_{(k)} = u_{(k)}\}$$

$$= \frac{u_{(n/2+1)} - u_{(n/2)}}{u_{(n/2+1)} - u_{(n/2-1)}} \quad ,$$

$$z_2 = P_0\{U_{(n/2+1)} \leq u_{(n/2+1)} | \text{all remaining } U_{(k)} = u_{(k)}\}$$

$$= \frac{[u_{(n/2+1)} - u_{(n/2-1)}]^2}{[u_{(n/2+2)} - u_{(n/2-1)}]^2} \quad ,$$

$$z_3 = \frac{[u_{(n/2+2)} - u_{(n/2-1)}]^3}{[u_{(n/2+2)} - u_{(n/2-2)}]^3} \quad ,$$

$$\vdots$$

$$z_n = u_{(n)}^n \quad .$$

For n odd, we define

$$z_1 = P_0\{U\left(\frac{n+1}{2}\right) \leq u\left(\frac{n+1}{2}\right) | \text{all other } U_{(k)} = u_{(k)}\}$$

$$= \frac{u\left(\frac{n+1}{2}\right) - u\left(\frac{n-1}{2}\right)}{u\left(\frac{n+3}{2}\right) - u\left(\frac{n-1}{2}\right)} ,$$

$$z_2 = \frac{\left[u\left(\frac{n+3}{2}\right) - u\left(\frac{n-1}{2}\right)\right]^2}{\left[u\left(\frac{n+3}{3}\right) - u\left(\frac{n-3}{2}\right)\right]^2} ,$$

$$z_3 = \frac{\left[u\left(\frac{n+3}{2}\right) - u\left(\frac{n-3}{2}\right)\right]^3}{\left[u\left(\frac{n+5}{2}\right) - u\left(\frac{n-3}{2}\right)\right]^3}$$

$$\vdots$$

$$z_n = u_{(n)}^n .$$

Finally, we define

$$L_k^{**} = 2 \operatorname{Min}\{Z_k, 1 - Z_k\}, \quad k = 1, 2, \ldots, n ,$$

and

$$Q_5 = -2 \sum_{k=1}^{n} \log L_k^{**} .$$

Small values of L_k^{**}, large values of Q_5, will be taken consistent with the alternative. For an observed q_5, the significance level, say ℓ_5, is

$$\ell_5 = P\{\chi^2(2n) \geq q_5\}$$

where $\chi^2(2n)$ is a chi-squared variable with $2n$ degrees of freedom.

Synthetic Sampling Study

Analytic studies of the sensitivity of goodness of fit test statistics are not very far advanced at present. It is difficult, from a practical point of view, to select the alternative hypothesis from all possible alternative hypotheses. There are also great mathematical difficulties connected with finding an exact, or even an approximation expression for the sensitivity of test statistics. This is particularly true for the type of test statistics defined in the preceding section, because the conditional levels ($L_{k|k+1}$, $L_{k+1|k}$, or L_k^{**}) are dependent random variables for most alternatives of interest. Even the theory of limiting distributions of statistics that are functions of dependent variables is not far advanced.

It is not difficult to obtain empirical approximations of distribution function (and sensitivities are distribution functions) with the aid of a high-speed computer. In a recent study by Shapiro, Wilk, and Chen (42) several of the test statistics defined in Chapter II were compared by comparing the empirical approximations of the distributions of significance levels. The study was designed to compare the W statistic (see Chapter II) with other tests of normality, thus in each case the null hypothesis was normal. Since the W is scale and origin invariant, the null hypothesis for this statistic was normal, mean and variance unknown. For the chi-squared statistic (CS^*), Kolmogorov-Smirnov (KS^*), Cramér-Von Mises (CM), weighted Cramér-Von Mises (WCM), and Durbin (D) it is necessary to specify the mean and variance. It is also necessary to specify the mean and variance for Q_3, Q_4, and Q_5 defined in the preceding section.

Accordingly, for the first set of alternatives (Table II), the mean and variance of the null hypothesis is taken as the known mean and variance of the actual alternative. For example, if the alternative is a chi-squared distribution with two degrees of freedom, the simple null is normal with mean two and variance four. This approach is of particular interest in the light of the transformations described in Chapter II designed to reduce the composite hypothesis of normality (mean and variance unknown) to a simple hypothesis; if one of the transformations is first applied to the data it will insure that the mean and variance of the alternative is nearly that of the null hypothesis, specifically mean zero and unit variance.

The values of β_1 and β_2 are the skewness and kurtois, respectively, of the corresponding alternative. Recall that a normal distribution, the null hypothesis, has $\beta_1 = 0$ and $\beta_2 = 3.0$.

Samples from the various alternatives were generated using the IBM subroutine RANDU as the basic imput (46). Obtaining samples was greatly simplified by the fact that all alternatives in Table II have distribution functions that are easy to invert. For example, if X is a chi-squared variable with two degrees of freedom, then

$$F(x) = 1 - e^{-x/2}, \quad x > 0,$$

and

$$x = -2 \log (1 - F(x)) .$$

Hence, to generate an observation x from this population, an observation u from a uniform on the unit interval is generated with RANDU, and then

$$x = -2 \log (1 - u) .$$

TABLE II

ALTERNATIVE DISTRIBUTIONS USED IN STUDY

Alternative	β_1, β_2
1. Uniform: $f_A(x) = 1,\ 0 \le x \le 1$ $= 0$, otherwise	$\beta_1 = 0.00$ $\beta_2 = 1.80$
2. Triangular: $f_A(x) = 2x,\ 0 \le x \le 1$ $= 0$, otherwise	$\beta_1 = -.57$ $\beta_2 = 2.40$
3. Chi-Squared, Two Degrees of Freedom: $f_A(x) = 0,\ x < 0$ $= 1/2\, e^{-x/2},\ x \ge 0$	$\beta_1 = 2.00$ $\beta_2 = 9.00$
4. Tukey ($\lambda = .7$): Tukey variates are defined by the transformation $Y = R^{\lambda} - (1-R)^{\lambda}$ where R is uniform on the unit interval	$\beta_1 = 0.00$ $\beta_2 = 1.92$
5. Tukey ($\lambda = 10$):	$\beta_1 = 0.00$ $\beta_2 = 5.38$
6. Laplace: $f_A(x) = 1/4\, e^{-1/2\lvert x\rvert},\ -\infty < x < \infty$	$\beta_1 = 0.00$ $\beta_2 = 6.00$
7. Weibull ($\lambda = 1, k = 0.5$): $f_A(x) = 0,\ x < 0$ $= \lambda k x^{k-1} e^{-\lambda x^k},\ x > 0$	$\beta_1 = 6.62$ $\beta_2 = 87.72$
8. Weibull ($\lambda = 1, k = 2$):	$\beta_1 = 0/63$ $\beta_2 = 3.25$
9. Logistic: $f_A(x) = \dfrac{2e^{2x}}{(1+e^{2x})^2},\ -\infty < x < \infty$	$\beta_1 = 0.63$ $\beta_2 = 3.25$
10. Cauchy: $f_A(x) = \dfrac{1}{\pi}\dfrac{1}{(1+x^2)},\ -\infty < x < \infty$	$\beta_1 = 0.00$ $\beta_2 = ----$

TABLE III

PERCENT SENSITIVITY AT THE 10 PERCENT LEVEL

OF SIGNIFICANCE; 200 SAMPLES FOR EACH

SAMPLE SIZE

	Test Statistic	Q_3			Q_5			KS*			CS*			CM			WCM			D		
	Sample Size	10	15	20	10	15	20	10	15	20	10	15	20	10	15	20	10	15	20	10	15	20
Alternative	Uniform	14	17	23	11	7	13	14	11	19	14	17	18	11	8	8	14	8	9	16	17	19
	Triangular	16	22	24	15	12	19	10	10	14	16	14	12	16	12	20	17	13	20	16	15	23
	Chi-Squared	38	51	57	43	46	56	32	30	37	81	43	43	20	23	33	23	26	41	31	39	56
	Tukey ($\lambda = .7$)	12	15	17	8	10	8	15	9	17	12	14	15	11	8	17	12	7	18	17	8	8
	Tukey ($\lambda = 10$)	73	87	93	78	89	96	31	45	63	27	83	81	23	33	54	27	29	61	72	77	87
	Laplace	19	24	26	23	26	33	13	7	13	14	17	25	12	7	15	36	20	37	23	21	24
	Weibull ($k = .5$)	94	100	100	98	100	100	58	65	100	94	97	99	57	73	96	63	77	99	89	99	100
	Weibull ($k = 2$)	12	13	19	10	12	14	12	14	15	11	15	12	10	11	14	11	11	15	10	7	11
	Logistic	13	14	21	15	16	18	9	6	9	12	12	8	10	9	8	13	9	9	13	4	10
	Cauchy	90	99	100	95	100	100	30	47	65	23	46	54	32	46	71	95	98	99	75	84	91

Samples of sizes $n = 10$, 15, and 20 were used in the study. For a typical computer run, that is, for one alternative, one sample size (say $n = 10$), and one test statistic (say Q_3), 200 samples of size 10 each were generated and 200 values of Q_3 computed. The percentage of the Q_3 values exceeding 28.41, the 90 percent quantile of a chi-squared distribution with 20 degrees of freedom, was then computed and recorded in Table III. Thus values given in Table III estimate the sensitivity at the 10 percent level of significance. Only percentages for Q_3 and Q_5 were actually computed in this study; all our percentages are those given by Shapiro, Wilk and Chen (42).

TABLE IV

PERCENT SENSITIVITY AT 5 PERCENT LEVEL OF SIGNIFICANCE; 200 SAMPLES OF SIZE 15

	Mean	.000	.000	.150	.300	.180	.360	.195	.390
	Std. Deviation	1.2	1.3	1.0	1.0	1.2	1.2	1.3	1.3
Test Statistic	Q_3	14	29	8	9	13	17	22	20
	Q_5	12	16	10	13	14	19	17	21
	KS^*	5	11	3	10	6	6	8	14
	CS^*	6	9	5	12	10	16	12	29
	CM	4	8	8	17	13	21	13	26
	WCM	11	15	9	17	20	29	26	38
	D	10	15	4	10	8	10	15	17

TABLE V

SUMS OF RANKS OF SENSITIVITIES; FIRST SET OF ALTERNATIVES

Test Statistic		Q_3	Q_5	KS*	CS*	CM	WCM	D
Sample Size	10	30.0	34.5	47.0	40.5	61.0	34.0	33.0
	15	17.5	27.0	53.5	32.5	57.0	51.0	41.5
	20	17.0	32.5	45.0	53.0	56.5	38.0	38.0

TABLE VI

SUMS OF RANKS OF SENSITIVITIES; SECOND SET OF ALTERNATIVES

Test Statistic		Q_3	Q_5	KS*	CS*	CM	WCM	D
Sample Size	15	27.0	20.0	51.5	38.0	32.5	14.0	41.0

The second set of alternatives are "misspecified" normal distributions used to study the effect of errors in the assumed values of the normal parameters in testing the simple hypothesis that the distribution is normal, mean zero and standard deviation one. The alternative parameter values are: $(\mu, \sigma) = (0, 1.2)$, $(0, 1.3)$, $(.15, 1.0)$, $(.18, 1.2)$, $(.195, 1.3)$, $(.3, 1.0)$, $(.36, 1.2)$, $(.39, 1.3)$. Results are given in Table IV.

To aid in summarizing the results of the synthetic sampling, ranks from one (best observed sensitivity) to seven (least observed sensitivity) were assigned for each sample size-alternative combination. These ranks were then summing over the ten alternative for the first study (Table V) and over the eight alternatives for the second study (Table VI). In case of ties ranks were averaged.

The results in this section were derived from sampling and are thus subject to sampling error. As a guide to accuracy, the standard deviation of any estimate of sensitivity is bounded by

$$\sqrt{\frac{(.5)(.5)}{200}} \approx .036 .$$

Even with the rather wide gauge of $\pm 2(.036) = \pm 7.2$ percentage points, Q_3, and Q_5 to a lesser extent, provide superior sensitivity for several of the alternatives considered. This is particularly true for alternatives with a long (heavy) tail compared to that of the normal distribution; these include the chi-squared, Tukey ($\lambda = 10$), Weibull ($k = 5$), and Cauchy alternatives.

It was mentioned in the preceding section that Q_3 generally provides better sensitivity when the mean of the alternative is smaller than that of the null, as opposed to the mean of the alternative greater than that of the null. To demonstrate this point, 200 samples of size 15 were generated from each of the alternatives normal $(\mu, \sigma) = (-.15, 1)$ and normal, $(\mu, \sigma) = (-.30, 1)$. With a null hypothesis of normal, $(\mu, \sigma) = (0, 1)$, the observed sensitivities were 16 and 25 percent, respectively, at the 5 percent level of significance. Compare these to the 8 and 9 percent given in Table IV for the alternatives normal, $(\mu, \sigma) = (.15, 1)$, and normal, $(\mu, \sigma) = (.30, 1)$.

CHAPTER VI

SUMMARY AND EXTENSIONS

In this study several test statistics were given for testing the simple goodness of fit hypothesis. Chapter III was concerned with constructing test statistics, referred to as methods of combination, for the special case of the general goodness of fit hypothesis obtained by considering significance levels of independent test statistics.

Properties of methods of combination were investigated in Chapter IV. The concept of a monotone method of combination was introduced, and relationships between monotonicity and the properties of unbiasedness and consistency of methods of combination were obtained. Specifically, if the p original test statistics are unbiased, then a monotone method of combination is an unbiased test statistic for the combined hypothesis problem; if the original test statistics are consistent, then a continuous (a random variable of the continuous type), monotone method of combination provides a consistent test statistic for a slightly restricted version of the combined hypothesis problem (see page 49). The joint integral transform method of combination is a continuous, monotone method of combination, as are Fisher's method, Pearson's method, the maximum level and the minimum level. However, the conditional integral transform methods W_3, W_4 (page 41), and W_5 (page 43) are not monotone methods of combination.

In the last section of Chapter IV it was shown that the exact slopes of the combined tests are $\sum_k c_k(\theta_k)$ for Fisher's method (established earlier by Littell (28)), $c_1(\theta_1) + c_2(\theta_2)$ for the joint integral transform when $p = 2$, $p \min_k c_k(\theta_k)$ for Pearson's method, and $\sum_k c_k(\theta_k)$ for the conditional integral transform method W_4 (with some restriction on the combined alternative). Littell also demonstrated that the slope of the method based on the minimum level is $\max_k c_k(\theta_k)$, and the slope of the method based on the maximum level is $p \min_k c_k(\theta_k)$. Thus based on the Bahadur slope criteria, for $p = 2$ and for the most general combined alternative, one should choose between Fisher's method and the joint integral transform method. For $p > 2$ perhaps Fisher's should be used. It is felt that the slope of the joint integral transform method is also the sum of the individual slopes for $p > 2$, however, it is not possible to prove this conjecture without the null distribution for general p.

Results of the synthetic sampling study given in Chapter IV clearly demonstrate that the conditional significance level approach to the simple goodness of fit problem is worthy of further consideration. However, more extensive synthetic sampling, with null hypotheses other than that of normality, would be necessary to draw general conclusions.

The composite goodness of fit problem definitely deserves further investigation. It is possible to use a technique, similar to the conditional significance level approach, to reduce the composite problem to a simple goodness of fit problem.

For example, suppose $X_1, X_2, \ldots, X_{n+1}, X_{n+2}$ are independent normally distributed random variables, each with the same unknown mean μ and unknown variance σ^2. Define

$$Y_k = \frac{X_k - \overline{X}}{\sqrt{\sum_{i=1}^{n+2} (X_i - \overline{X})^2}} \quad , \quad k = 1, 2, \ldots, n$$

where

$$\overline{X} = \frac{1}{(n+2)} \left(\sum_{i=1}^{n+2} X_i \right),$$

It is shown in the appendix that the joint density of $Y_1, Y_2, \ldots, Y_n$ does not depend on μ and σ^2, however, the Y's are dependent random variables.

Suppose we apply the technique described in Chapter III (pages 40-42) to obtain mutually independent uniform variables. That is, define $U_1, U_2, \ldots, U_n$ by

$$U_{n-r} = F_{Y_{n-r} | Y_1, \ldots, Y_{n-r-1}}(Y_{n-r} | Y_1, \ldots, Y_{n-r-1}), \quad r = 0, 1, 2, \ldots, n\text{-}1$$

where $F_{Y_{n-r} | Y_1, \ldots, Y_{n-r-1}}(\cdot|\cdot)$ is the conditional distribution function of Y_{n-r} given $Y_1, Y_2, \ldots, Y_{n-r-1}$. It is shown in the appendix that U_{n-r} will be

$$U_{n-r} = \frac{1}{2} - \frac{1}{2}\, \beta(Z^2_{n-r} ; \frac{1}{2}, \frac{r+1}{2}), \quad Z_{n-r} \leq 0 ,$$

$$U_{n-r} = \frac{1}{2} + \frac{1}{2}\, \beta(Z^2_{n-r} ; \frac{1}{2}, \frac{r+1}{2}), \quad Z_{n-r} > 0 ,$$

where

$$Z_{n-r} = \frac{Y_{n-r} + \frac{1}{(r+3)} \sum_{i=1}^{n-r-1} Y_i}{\sqrt{\left(\frac{r+2}{r+3}\right)\left\{1 - \left(\sum_{i=1}^{n-r-1} Y_i^2\right) - \frac{1}{(r+3)}\left(\sum_{i=1}^{n-r-1} Y_i\right)^2\right\}}},$$

and $\beta(\cdot\,;\alpha,\beta)$ is the incomplete beta function (see appendix).

Therefore, the composite hypothesis concerning $X_1, X_2, \ldots, X_{n+1}, X_{n+2}$ can be tested as a simple hypothesis concerning the independent (under the null hypothesis) random variables $U_1, U_2, \ldots, U_n$. The appealing feature of this approach is that we essentially replace the unknown parameter (μ, σ^2) by its minimal sufficient statistic $(\overline{X}, \Sigma(X_i - \overline{X})^2)$. No attempt has been made to study the sensitivity of this approach.

A SELECTED BIBLIOGRAPHY

(1) Anderson, T. W. and Darling, D. A. "A Test of Goodness of Fit." Journal of the American Statistical Association. Vol. 49 (1954) 765-799.

(2) Bahadur, R. R. "Stochastic Comparisons of Tests." Annals of Mathematical Statistics. Vol. 31 (1960) 276-295.

(3) Bahadur, R. R. "Rates of Convergence of Estimates and Test Statistics." Annals of Mathematical Statistics. Vol. 38 (1967) 303-324.

(4) Birnbaum, A. "Combining Independent Tests of Significance." Journal of the American Statistical Association. Vol. 49 (1954) 559-575.

(5) Birnbaum, Z. W. "Numerical Tabulation of the Distribution of Kolmogorov Statistic for Finite Sample Size." Journal of the American Statistical Association. Vol. 47 (1952) 425-441.

(6) Chapman, D. G. "A Comparative Study of Several One-Sided Goodness-of-Fit Tests." Annals of Mathematical Statistics. Vol. 29, Pt. 2 (1958) 655-673.

(7) Chernoff, H. and Lehmann, E. L. "The Use of Maximum Likelihood Estimates in χ^2 Tests for Goodness of Fit." Annals of Mathematical Statistics. Vol. 25 (1954) 579-586.

(8) Cochran, W. G. "The χ^2 Test of Goodness of Fit." Annals of Mathematical Statistics. Vol. 23 (1952) 315-345.

(9) Cramér, H. Mathematical Methods of Statistics. Princeton University Press, Princeton, N. J. (1946).

(10) Darling, D. A. "The Kolmogorov-Smirnov, Cramér-Von Mises Tests." Annals of Mathematical Statistics. Vol. 28 (1957) 823-838.

(11) David, H. T. "Goodness of Fit." International Encyclopedia of Social Sciences. (1968) 199-207.

(12) Durbin, J. "Some Methods of Constructing Exact Tests." Biometrika. Vol. 48 (1961) 41-55.

(13) Finley, R. D. "A Theory of Significance Testing." Unpublished Ph.D. Thesis (1968). Oklahoma State University.

(14) Fisher, R. A. "The Conditions Under Which χ^2 Measures the Discrepancy Between Observation and Hypothesis." Journal of the Royal Statistical Society. Vol. 87 (1924) 442-450.

(15) Fisher, R. A. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh and London., (1932).

(16) Fisz, M. Probability Theory and Mathematical Statistics. John Wiley and Sons, Inc., New York, N.Y., (1963).

(17) Fraser, D. A. S, Nonparametric Methods in Statistics. John Wiley and Sons, Inc., New York, N.Y., (1957).

(18) Good, I. J. "On the Weighted Combination of Significance Tests." Journal of the Royal Statistical Society, Series B. Vol. 17, (1955) 264-265.

(19) Graybill, F. A. An Introduction to Linear Statistical Models. McGraw-Hill Book Company, Inc., New York, N.Y., (1961).

(20) Keats, J. A. and Lord, F. M. "A Theoretical Distribution for Mental Test Scores." Psychometrika. Vol. 27 (1962) 59-72.

(21) Kempthorne, O. "The Classical Problem of Inference - Goodness of Fit." Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. (1966) 235-249.

(22) Kendall, M. G. and Stuart, A. The Advanced Theory of Statistics Vol. 1, Hafner Publishing Company, New York, N.Y., (1958).

(23) Kolmogorov, A. N. "Sulla Determinazione Empirica di une Legge di Distribuzione." Giornale. Vol. 4 (1933) 83-99.

(24) Lehmann, E. L. "Consistency and Unbiasedness of Certain Non-Parametric Tests." Annals of Mathematical Statistics. Vol. 24 (1951) 165-184.

(25) Lehmann, E. L. "The Power of Rank Tests." Annals of Mathematical Statistics. Vol. 24 (1953) 23-32.

(26) Lehmann, E. L. Testing Statistical Hypotheses. John Wiley and Sons, Inc., New York, N.Y., (1957).

(27) Liptak, T. "On the Combination of Independent Tests." Magyar Tvd. Akad. Mat. Kutató Int. Közl. Vol. 3 (1958) 171-197.

(28) Littell, R. C. "Properties of Bahadur Efficiency." Unpublished Ph.D. Thesis (1970). Oklahoma State University.

(29) Littell, R. C. and Folks, J. L. "On the Relation Between Bahadur Efficiency and Power." Annals of Mathematical Statistics. Vol. 41 (1970) 2135-2136.

(30) Marshall, A. W. "The Small Sample Distribution of nw_n^2." Annals of Mathematical Statistics. Vol. 29 (1958) 307-309.

(31) Massey, F. J. "The Kolmogorov-Smirnov Test for Goodness of Fit." Journal of the American Statistical Association. Vol. 46 (1951) 68-78.

(32) Moore, B. J. "On A Theory of Significance Testing." Unpublished Ph.D. Thesis (1969). Oklahoma State University.

(33) Neymann, J. "Contribution to the Theory of the χ^2 Test." Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability. (1949) 239-273.

(34) Pearson, E. S. "The Probability Integral Transformation for Testing Goodness of Fit and Combining Independent Tests of Significance." Biometrika. Vol. 30 (1938) 134-148.

(35) Pearson, Karl. "On the Criterion That a Given System of Deviations From the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen From Random Sampling." Philosophical Magazine 5th Series. Vol. 50 (1900) 157-175.

(36) Prohorov, Y. "Some Characterization Theorems in Statistics." Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. (1966) 341-349.

(37) Pyke, R. "Spacings." Journal of the Royal Statistical Society, Series B. Vol. 27 (1965) 395-449.

(38) Rao, C. R. Advanced Statistical Methods in Biometric Research. John Wiley and Sons, New York, N.Y., (1952).

(39) Sarkadi, K. "On Testing for Normality." Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. (1966) 373-387.

(40) Savage, I. R. "Nonparametric Statistics: A Personal Review." Sankhya. Vol. 31 (1969) 107-144.

(41) Shapiro, S. S. and Wilk, M. B. "An Analysis of Variance Test of Normality (Complete Samples)." Biometrika. Vol. 52 (1965) 591-611.

(42) Shapiro, S. S., Wilk, M. B., and Chen, H. J. "A Comparative Study of Various Tests for Normality." Journal of the American Statistical Society. Vol. 63 (1968) 1343-1372.

(43) Slutsky, E. E. "Über Stochastische Asymptoten und Grenzwerte." Metron. Vol. 5 (1925) 3.

(44) Smirnov, N. V. "On the Estimation of the Discrepancy Between Empirical Curves of Distribution for Two Independent Samples." Bulletin Mathematique. Vol. 2 (1939) 3-26.

(45) Suzuki, G. "Kolmogorov-Smirnov Tests of Fit Based on Some General Bounds." Journal of the American Statistical Association, Vol. 63 (1968) 919-924.

(46) System/360 Scientific Subroutine Package (360 A-CM-03X) Version III Programmer's Manual. International Business Machines Corporation, White Plains, N. Y., (1966).

(47) van Zwet, W. R. and Oosterhoff, J. "On the Combination of Independent Test Statistics." Vol. 40. (1969) 659-680.

(48) Wallis, W. Allen. "Compounding Probabilities from Independent Significance Tests." Econometrica. Vol. 10 (1942) 229-248.

(49) Watson, G. S. "The χ^2 Goodness-of-Fit Test for Normal Distributions." Biometrika. Vol. 44 (1957) 336-348.

(50) Weiss, L. "Limiting Distribution of Homogeneous Functions of Sample Spacings." Annals of Mathematical Statistics. Vol. 39 (1958) 310-312.

(51) Wilkinson, B. "A Statistical Consideration in Psychological Research." Psychological Bulletin. Vol. 48 (1951) 156-157.

APPENDIX

The purpose of this appendix is to provide details of transformations, given in Chapter VI, to reduce the composite hypothesis of normality, mean μ and variance σ^2 unknown, to a simple hypothesis. Let $X_1, X_2, \ldots, X_{n+2}$ denote independent normally distributed random variables, each with the same (unknown) mean μ and variance σ^2. Define

$$W_1 = X_{n+1} + X_{n+2}$$

and

$$W_2 = (X_{n+1} - W_1/2)^2 + (X_{n+2} - W_1/2)^2 ,$$

then W_1 and W_2 are independent with respective density functions

$$f(w_1) = \frac{1}{\sqrt{2\pi}\,\sqrt{2}\,\sigma} \exp - \frac{1}{4\sigma^2} (w_1 - 2\mu)^2 , \quad -\infty \leq w_1 < \infty ,$$

and

$$f(w_2) = \frac{1}{\sqrt{2\pi}\,\sigma} w_2^{-1/2} \exp - \frac{w_2}{2\sigma^2} , \quad w_2 > 0$$

$$= 0 , \text{ otherwise} .$$

The joint density of $X_1, X_2, \ldots, X_n, W_1, W_2$ is

$$f(x_1, \ldots, x_n, w_1, w_2) = \frac{w_2^{-1/2} \exp - \frac{1}{2\sigma^2} \left[\frac{1}{2} (w_1 - 2\mu)^2 + w_2 + \sum_{k=1}^{n} (x_k - u)^2 \right]}{\pi^{(n+2)/2}\, 2^{(n+3)/2}\, \sigma^{n+2}} ,$$

$$-\infty < x_k < \infty \ ,$$
$$-\infty < w_1 < \infty \ ,$$
$$0 < w_2 < \infty \ .$$

Make the change of variables

$$T_1 = W_1 + \sum_{k=1}^{n} X_k \ ,$$

$$T_2 = W_2 + \frac{(T_1 - \sum_{k=1}^{n} X_k)^2}{2} + \sum_{k=1}^{n} X_k^2 - \frac{T_1^2}{n+2} \ ,$$

$$Y_k = \frac{X_k - \frac{1}{(n+2)} T_1}{\sqrt{T_2}} \ , \quad k = 1, 2, \ldots, n \ ,$$

then the joint density of $Y_1, \ldots, Y_n, T_1, T_2$ is

$f(y_1, \ldots, y_n, t_1, t_2)$

$$= \frac{t_2^{n+1/2}[1 - \vec{y}_n' B_n \vec{y}_n]^{-1/2} \exp - \frac{1}{2\sigma^2} [\frac{1}{(n+2)} (t_1 - (n+2)\mu)^2 + t_2]}{\pi^{(n+2)/2} \, 2^{(n+3)/2} \, \sigma^{n+2}} \ ,$$

$$0 < \vec{y}_n' B_n \vec{y}_n < 1 \ ,$$
$$-\infty < t_1 < \infty \ ,$$
$$0 < t_2 < \infty \ ,$$

where

$$\vec{y}_n = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

$$B_n = I_n + \frac{1}{2} J_n J_n' ,$$

I_n is the n dimensional identity matrix and J_n is an $n \times 1$ vector of ones.

Note that

$$T_1 = \sum_{k=1}^{n+2} X_k ,$$

$$T_2 = \sum_{k=1}^{n+2} (X_k - \overline{X})^2$$

where

$$\overline{X} = \frac{1}{(n+2)} T_1 .$$

Therefore the joint density of T_1 and T_2 is

$$f(t_1, t_2) = \frac{t_2^{n+1/2} \exp - \frac{1}{2\sigma^2} [\frac{1}{(n+2)} (t_1 - (n+2)\mu)^2 + t_2]}{\sqrt{2\pi (n+2) \sigma^2}\, \Gamma\left(\frac{n+1}{2}\right) 2^{(n+1)/2} \sigma^{n+1}}$$

and the conditional density of $Y_1, Y_2, \ldots, Y_n$ given T_1, T_2 is

$$f(y_1, \ldots, y_n | t_1, t_2) = \frac{\Gamma \frac{n+1}{2} \sqrt{n+2}}{\sqrt{2}\, \pi^{(n+1)/2}} [1 - \vec{y}_n' B_n \vec{y}_n]^{-1/2}, \quad 0 < \vec{y}_n' B_n \vec{y}_n < 1 .$$

Since $f(y_1, \ldots, y_n | t_1, t_2)$ does not depend on t_1 and t_2, this is just the joint density of $Y_1, Y_2, \ldots, Y_n$.

In the remainder of this appendix the following notation will be used:

(i) I_{n-r} will denote the $n-r$ dimensional identity, $r = 0, 1, \ldots, n-1$;

(ii) J_{n-r} will denote an $n-r$ dimensional vector of ones, $r = 0, 1, \ldots, n-1$;

(iii) B_{n-r} will denote the matrix

$$B_{n-r} = I_{n-r} + \frac{1}{(r+2)} J_{n-r} J'_{n-r}, \quad r = 0, 1, \ldots, n-1 ;$$

(iv) $\vec{y}_{n-r}$ will denote the vector

$$\vec{y}_{n-r} = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_{n-r} \end{bmatrix}, \quad r = 0, 1, \ldots, n-1 .$$

Now assume that, for $r = 0, 1, \ldots, k$, the joint density of $Y_1, \ldots, Y_{n-r}$ is

$$f(y_1, \ldots, y_{n-r}) = C_{n-r} [1 - \vec{y}'_{n-r} B_{n-r} \vec{y}_{n-r}]^{(r+1)/2-1} ,$$

$$0 < \vec{y}'_{n-r} B_{n-r} \vec{y}_{n-r} < 1 .$$

where C_{n-r} is a constant such that the density integrates to one. Now

$$\vec{y}'_k B_{n-k} \vec{y}_{n-k} = \sum_{i=1}^{n-k} y_i^2 + \frac{1}{k+2} \sum_{j=1}^{n-k} \sum_{i=1}^{n-k} y_i y_j$$

$$= \left(\frac{k+3}{k+2}\right) y_{n-k}^2 + 2\, y_{n-k} \left(\frac{1}{k+2}\right) \vec{y}'_{n-k-1} J_{n-k+1}$$

$$+ \vec{y}'_{n-k-1} \{I_{n-k-1} + \left(\frac{1}{k+2}\right) J_{n-k-1} J'_{n-k-1}\} \vec{y}_{n-k-1}$$

$$= \left(\frac{k+3}{k+2}\right)\{y_{n-r} + \frac{1}{k+3}\vec{y}'_{n-k-1} J_{n-k-1}\}^2$$
$$+ \vec{y}'_{n-k-1} B_{n-k-1} \vec{y}_{n-k-1}$$

Let

$$m_r = \frac{1}{r+2}\vec{y}_{n-r} J_{n-r}$$

and

$$s_r^2 = \vec{y}'_{n-r} B_{n-r} \vec{y}_{n-r} ,$$

then

$$f(y_1, \ldots, y_{n-k-1}) = C_{n-k}[1 - s_{k+1}^2]^{(k+2)/-1}\sqrt{\frac{k+2}{k+3}}\int_{-1}^{1}(1-t^2)^{(k+1)/2-1}dt$$
$$= C_{n-k-1}[1 - \vec{y}'_{n-(k+1)} B_{n-(k+1)} \vec{y}_{n-(k+1)}]^{\frac{(k+1)+1}{2} - 1} .$$

Since the original assumption is true for $r = 0$, it is true for $r = 1, 2, \ldots, n-1$ by induction.

It can be shown that

$$\int_{-1}^{1}(1-t^2)^{\frac{r+1}{2}-1} dt = \frac{\sqrt{\pi}\,\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r+2}{2}\right)}$$

so

$$f(y_{n-r} | y_1, \ldots, y_{n-r-1})$$
$$= \frac{\Gamma\left(\frac{r+2}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{r+1}{2}\right)}\sqrt{\frac{r+3}{r+2}}(1 - s_{r+1}^2)^{-1/2}\left[1 - \frac{(y_{n-r} + m_{r+1})^2}{\left(\frac{r+2}{r+3}\right)(1 - s_{r+1}^2)}\right] ,$$

$$- m_{r+1} - \sqrt{\frac{r+2}{r+3}}\sqrt{1 - s_{r+1}^2} < y_{n-r} < -m_{r+1} + \sqrt{\frac{r+2}{r+3}}\sqrt{1 - s_{r+1}^2}$$

Therefore

$$F(y_{n-r}|y_1,\ldots,y_{n-r-1}) = \frac{\Gamma\left(\frac{r+2}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{r+1}{2}\right)} \int_{-1}^{z_{n-r}} (1-t^2)^{\frac{r+1}{2}-1}\,dt\ ,$$

$$r = 0, 1, 2, \ldots, n-1$$

where

$$z_{n-r} = \frac{y_{n-r} + m_{r+1}}{\sqrt{\frac{r+2}{r+3}}\ \sqrt{1 - s_{r+1}^2}}\ .$$

If we make the change of variables in integration $v = t^2$ then

$$F(y_{n-r}|y_1,\ldots,y_{n-r-1}) = \frac{1}{2} - \frac{1}{2}\,\beta(z_{n-r}^2\,; \frac{1}{2}, \frac{r+1}{2})\ ,\quad z_{n-r} < 0\ ,$$

$$F(y_{n-r}|y_1,\ldots,y_{n-r-1}) = \frac{1}{2} + \frac{1}{2}\,\beta(z_{n-r}^2\,; \frac{1}{2}, \frac{r+1}{2})\ ,\quad z_{n-r} > 0\ ,$$

where $\beta(\cdot\,;\alpha,\beta)$ is the incomplete beta function; that is

$$\beta(x\,;\alpha,\beta) = \int_0^x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, t^{\alpha-1}(1-t)^{\beta-1}\,dt\ ,$$

$$0 \le x \le 1,\ \alpha > -1,\ \beta > -1\ ,$$

VITA

Charles Cleveland Stewart

Candidate for the Degree of

Doctor of Philosophy

Thesis: A JOINT INTEGRAL TRANSFORMATION APPROACH TO GOODNESS OF FIT

Major Field: Statistics

Biographical:

Personal Data: Born in Woodward, Oklahoma, September 12, 1941, the son of Mr. and Mrs. Orval F. Stewart.

Education: Graduated from Woodward High School, Woodward, Oklahoma, May, 1959; received the Bachelor of Science degree from Oklahoma State University in May, 1963, with a major in Mathematics; received the Master of Science degree from Oklahoma State University in May, 1966, as a National Defense Educational Act Fellow, with a major in Mathematics and Statistics; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in July, 1971.

Professional Experience: Scientific Staff Member, Sandia Corporation, Summer 1964; Military Analyst, Joint Task Force Two, 1967-1969; Systems Analyst, Nuclear Training Directorate, 1969; Graduate Teaching Assistant, Department of Mathematics and Statistics, Oklahoma State University, 1969-1971.