ON GOODNESS-OF-FIT TESTS

BASED ON RAO-BLACKWELL DISTRIBUTION FUNCTION ESTIMATORS

by

Federico Jorge O'Reilly

Institute of Statistics
Mimeograph Series No. 788
Raleigh - December 1971
TABLE OF CONTENTS

1. INTRODUCTION AND SUMMARY .............................................. 1

2. REVIEW OF LITERATURE .................................................... 4

3. UNIFORM STRONG CONVERGENCE OF RAO-BLACKWELL DISTRIBUTION
   FUNCTION ESTIMATORS .................................................. 6
   3.1 Introduction ......................................................... 6
   3.2 A Glivenko-Cantelli Type Result .................................. 8
   3.3 Strong Consistency of Quantile Estimators ................. 11

4. GOODNESS-OF-FIT TESTS .................................................. 16
   4.1 Statement of the Problem .......................................... 16
   4.2 A Distribution-Free Test ......................................... 17
   4.3 The Conditional Probability Integral
       Transformation ..................................................... 24

5. CHI-SQUARE TYPE STATISTICS .......................................... 37
   5.1 Simple Goodness-of-Fit Problem .................................. 37
   5.2 Composite Goodness-of-Fit Problem ........................... 41

6. APPLICATIONS ............................................................. 45
   6.1 Test for Normality, \( \mu, \sigma^2 \) Unknown .................. 45
   6.2 Test for Exponentiality ........................................... 47
   6.3 Multivariate Normal Test ......................................... 48
   6.4 Testing the Fit of a Regression Model ....................... 50

7. LIST OF REFERENCES ..................................................... 54
1. INTRODUCTION AND SUMMARY

Goodness-of-fit problems have long received considerable attention due to their importance in practice. Simple goodness-of-fit problems, i.e., those in which the null hypothesis consists of a single distribution function, have been widely studied for the univariate case. Distribution-free and asymptotically distribution-free tests are common in the literature.

The concept of distribution-freeness has been generalized here for the composite goodness-of-fit problem by requiring invariance of the distribution of the test statistic under choices of distribution functions in the null class. The tests so far proposed are clearly generalizations of the simple goodness-of-fit tests using maximum likelihood estimation and more recently minimum variance unbiased estimation.

Under finite sample theory some of the available tests for composite goodness-of-fit problems have been shown to be distribution-free, however, the distribution of the test statistic fails to be invariant for different choices of the null hypothesis class. This fact provided motivation for the present work.

In this thesis, the notion of distribution-freeness is analysed for composite goodness-of-fit problems. Sufficient conditions are established for a problem to admit a distribution-free test, and then the notion of class-freeness is introduced as the invariance of the distribution of a test statistic, not only under choices of particular
members of the null class to be tested, but under the choice of null classes, also.

Rao-Blackwell distribution function estimators are used throughout this work. In Chapter 3, the asymptotic behaviour of Rao-Blackwell estimators is studied. A Glivenko-Cantelli type result is shown for the Rao-Blackwell distribution function estimators, from which strong consistency of quantile estimators is derived as an example of immediate application.

In Chapter 4, the notions of distribution-freeness and class-freeness are studied. Sufficient conditions for a generalized Kolmogorov-Smirnov statistic to be distribution-free are obtained. A generalization of the probability integral transformation is given, which under certain conditions, reflected in the absolute continuity of the Rao-Blackwell distribution estimators of the null class to be tested, allows a subset of the sample to be transformed into independently distributed random variables on the unit interval.

In Chapter 5, the chi-square statistic is proposed as a test statistic for the sample obtained by transforming the original observations. In this case, it is shown how the test can also be carried out with the beginning sample (actually a subset of it), with the difference that not only the cells are selected at random, but for each observation of the sample, a different set of random cells must be selected. The distribution of the chi-square type statistics proposed is identified as that of the exact chi-square statistic for testing a simple goodness-of-fit problem with equiprobable cells.
In Chapter 6, some direct applications are given. Univariate normal and exponential fit problems are considered. One multivariate problem is analysed and a test for the fit of a regression model is given.
2. REVIEW OF LITERATURE

The problem of testing the goodness-of-fit of a given distribution or a given class of distribution functions by means of an independent sample is an old one. Perhaps the earliest test developed for this purpose has been the chi-square goodness-of-fit test (usually denoted $X^2$), which has been widely used. Cochran [4], gives an excellent exposition of this test.

Later, the Kolmogorov-Smirnov and Cramér-von Mises statistics were introduced; these are based on stronger theoretical aspects and more sophisticated "distances" between distribution functions. Darling [5], discusses these tests and their properties.

In dealing with the problem of testing a class of distributions, i.e., a composite goodness-of-fit problem, under fixed selection of cells, and under certain regularity conditions, a well known result is that the $X^2$ statistic, computed with maximum likelihood estimators (or minimum $X^2$ estimators) is asymptotically distribution-free.

If the composite problem has a null class whose members are identifiable by location and scale parameters $\theta_1, \theta_2$; with proper estimates $\hat{\theta}_1, \hat{\theta}_2$, the Kolmogorov-Smirnov statistic computed with the empirical distribution $F_n(x)$ and $F(x; \hat{\theta}_1, \hat{\theta}_2)$ is distribution-free (see David and Johnson [6]). Lilliefors, [11] and [12] simulated the distribution of the statistic under the null hypothesis for the normal and exponential case.

Srinivasan [21], has shown that the Kolmogorov-Smirnov statistic, constructed with the minimum variance unbiased estimator (MVUE) $\hat{F}$ and
the usual empirical distribution $F_n$, is distribution-free for testing normality, and also for testing exponentiality.

Watson [22] and [23] has shown that for a composite goodness-of-fit problem, the $X^2$ statistic, computed with a random selection of cells is asymptotically distribution-free in the case of location and scale parameters, and, in general, he identifies the asymptotic distribution as that of:

$$\sum_{i=1}^{k-m-1} Z_i^2 + \sum_{i=k-m}^{k} \lambda_i Z_i^2,$$

where $Z_1, \ldots, Z_k$ are NID(0, 1), and $\lambda_{k-m}, \ldots, \lambda_k$ are some eigen values in $(0, 1)$ which he identifies.

Moore [15] has extended Watson's results to the multivariate case. Like Watson, he uses ML estimators and his assumptions contain the "regularity" assumptions for an estimation problem.

In this thesis, the papers used for reference in minimum variance unbiased distribution estimation are those of Sathe and Varde [18], for the univariate case, and Ghurye and Olkin [7], for the multivariate normal and related distributions.
3. UNIFORM STRONG CONVERGENCE OF RAO-BLACKWELL
DISTRIBUTION FUNCTION ESTIMATORS

3.1 Introduction

Let \((\mathbb{R}^\infty, \mathcal{F}^\infty, P^\infty)\) be the probability space corresponding to a sequence of independent observations \(Y_1, \ldots, Y_n, \ldots\) of the random variable \(Y\) distributed on the Borel line \((\mathbb{R}, \mathfrak{B})\) according to the probability measure \(P\) whose associated distribution function is \(F\). This probability space will be referred to as the independent sampling model.

For each \(n \geq 1\) let \(U_n : \mathbb{R}^\infty \to \mathbb{R}^n\) be the vector of order statistics of the first \(n\) observations. Define for each \(x \in \mathbb{R}\),

\[\delta_x : \mathbb{R}^\infty \to \{0, 1\},\]

by

\[\delta_x = \begin{cases} 1, & \text{if the first coordinate of the sample point is } \leq x; \\ 0, & \text{elsewhere}. \end{cases}\]

Let \(\mathcal{C}^n\) be the sub \(\sigma\)-algebra of \(\mathcal{F}^\infty\) induced by \(U_n\). The empirical distribution function evaluated at \(x\), computed from the first \(n\) terms of the sample is:

\[F_n(x) = \mathbb{E}(\delta_x | \mathcal{C}^n).\]

The Glivenko-Cantelli Lemma says that with probability 1,

\[F_n(x) \to F(x) \text{ uniformly in } x, \text{ i.e.,} \]

\[\sup_{x \in \mathbb{R}} \left| \mathbb{E}(\delta_x | \mathcal{C}^n) - \mathbb{E}(\delta_x) \right| \to 0 \text{ a.s. as } n \to \infty.\]
For each $n \geq 1$, let $T_n: \mathcal{F}^\infty \rightarrow \mathbb{R}^n$, ($k_n$ some integer $\leq n$), be such that:

3.1.1 \hspace{1cm} T_n \text{ is } \mathcal{F}^\infty - \mathcal{B}^n \text{ measurable,}

3.1.2 \hspace{1cm} T_n \text{ induces the } \sigma\text{-algebra } \mathcal{B}^n \subset \mathcal{C}^n \text{ (symmetry)},

3.1.3 \hspace{1cm} \mathcal{B}^{n+1} \subset \mathcal{B}^n \sigma (Y_{n+1}), \text{ where } \mathcal{B}^n \sigma (Y_{n+1}) \text{ is the compound } \sigma\text{-algebra of } \mathcal{B}^n \text{ and that induced by the } (n+1)\text{st observation (sequentiality}.}

Definition

A sequence of statistics $(T_n)_{n \geq 1}$ defined on $\mathcal{F}^\infty$, $\mathcal{B}^\infty$, $\mathcal{C}^\infty$, for which 3.1.1 through 3.1.3 is satisfied will be said to be a sequence of symmetric sequential statistics.

Berk [1] has shown that for a sequence of symmetric sequential statistics,

3.1.4 \hspace{1cm} E(Z|\mathcal{B}^n) = E(Z|\mathcal{B}^{n+1} \ldots) \text{ a.s.}

for any $Z$ that is $\sigma(Y_1, \ldots, Y_k)$ measurable ($k < n$), and integrable $\mathcal{F}^\infty$.

With 3.1.4 and the fact that $(\mathcal{B}^{n+1} \ldots)_{n \geq 1}$ is a monotonically contracting sequence of $\sigma$-algebras, he shows that $E(Z|\mathcal{B}^n) \rightarrow E(Z|\mathcal{F}^\infty)$ a.s., where $\mathcal{F}^\infty$ is the tail $\sigma$-algebra of the sequence $(\mathcal{B}^n)_{n \geq 1}$, which by virtue of the Hewitt-Savage, 0-1 law, is a.s. trivial, i.e., $E(Z|\mathcal{B}^n) \rightarrow E(Z)$ a.s.

Berk's result can be summarized as follows:
"Every Rao-Blackwell estimator based on a symmetric sequential statistics converges strongly to its expectation, whenever this is finite."

Where, for any r.v. Z, the Rao-Blackwell estimator of E(Z) based on a sufficient σ-algebra \( \mathcal{F}^n \) is \( E(Z | \mathcal{F}^n) \).

### 3.2 A Glivenko-Cantelli Type Result

Loeve [13], p. 409, states the following result:

**Result 3.2.1**

Let \((\mathcal{G}^n)_{n \geq 1}\) be a sequence of monotonic sub σ-algebras and let \((Z_m)_{m \geq 1}\) be a sequence of r.v.'s converging to Z a.s. such that

\[
\sup_{m \geq 1} |Z_m| \in L_r
\]

for some \( r \geq 1 \); then,

\[
E(Z_m | \mathcal{G}^n) \to E(Z | \mathcal{G}^\infty) \quad \text{a.s.}
\]

as \( m, n \to \infty \), where \( \mathcal{G}^\infty \) is the tail σ-algebra of the sequence \((\mathcal{G}^n)_{n \geq 1}\).

**Theorem 3.2**

In the independent sampling model, let \((\mathcal{G}^n)_{n \geq 1}\) be sub σ-algebras induced by symmetric sequential statistics. Then,

\[
\sup_{x \in \mathbb{R}} |E(\delta_x | \mathcal{G}^n) - E(\delta_x)| \to 0, \quad \text{a.s.}
\]
Proof

First note that for \((C^n)_{n \geq 1}\), the sequence of sub \(\sigma\)-algebras induced by the order statistics vectors,

\[ E(\delta^i \mid C^n) = (1/n) \sum_{i=1}^{n} \delta^i, \]

where

\[
\delta^i : \mathbb{R}^\infty \rightarrow \{0, 1\}
\]

is defined for each \(x \in \mathbb{R}\), by:

\[
\delta^i_x = \begin{cases} 
1, & \text{if the } i^{th} \text{ coordinate of the sample point is } \leq x, \\
0, & \text{elsewhere.}
\end{cases}
\]

Also, note that:

\[ E(\delta^i \mid C^n) = E(\delta^i \mid C^n) \text{ for } i = 1, \ldots, n, \]

therefore, since,

\[ E(\delta^i \mid B^n) = E[E(\delta^i \mid C^n) \mid B^n], \]

it follows that

\[ E(\delta^i \mid B^n) = E[E(\delta^i \mid C^n) \mid B^n] \text{ for } i = 1, \ldots, n; \]

from which,

\[ E(\delta^i \mid B^n) \]

can be thus expressed as:
\[(1/k) \sum_{i=1}^{k} E(\delta_{x}^{i} | \mathcal{F}^{n}), \]

for any \( k = 1, \ldots, n, \) i.e.,

\[E(\delta_{x} | \mathcal{F}^{n}) = E[E(\delta_{x} | \mathcal{C}^{k}) | \mathcal{F}^{n}], \quad 1 \leq k \leq n.\]

Denote by \( Z_{n} \) the \( \mathcal{C}^{n-1} \) measurable random variable

\[Z_{n} = \sup_{x \in \mathbb{R}} |E(\delta_{x} | \mathcal{C}^{n-1}) - E(\delta_{x})|.\]

The sequence \( (Z_{n})_{n \geq 2} \) is non-negative, uniformly bounded by 1 and a.s. convergent to 0, therefore,

\[\sup_{n \geq 2} Z_{n} \in L_{r}, \quad \forall r \geq 1.\]

Also, note that

\[E(Z_{n} | \mathcal{F}^{n}) = \sup_{x \in \mathbb{R}} |E(\delta_{x} | \mathcal{F}^{n}) - E(\delta_{x})|,\]

so it will suffice to show that:

\[E(Z_{n} | \mathcal{F}^{n}) \to 0 \text{ a.s.}\]

Because of 3.1.4,

\[E(Z_{n} | \mathcal{F}^{n}) = E(Z_{n} | \mathcal{G}^{n}) \text{ a.s.,}\]

where

\[\mathcal{G}^{n} = \mathcal{F}^{n} \mathcal{G}^{n+1} \ldots\]

is monotonic with a.s. trivial limit; hence, by Result 3.2.1,

\[E(Z_{n} | \mathcal{F}^{n}) \to 0 \text{ a.s.}\]
The random variable $E(\delta_x|\theta^n)$, will be denoted $\tilde{T}_n(x)$, and referred to as the Rao-Blackwell distribution function estimator obtained by conditioning on $\theta^n$ or $T_n$.

Note:

Multivariate generalizations of the Glivenko-Cantelli Lemma have been given in the literature. See, for example, Ranga Rao [16], Wolfowitz [24], [25], and Blum [3]. It is of interest to note that Theorem 3.2 would generalize also, whenever the conditions of symmetry and sequentiality are met. Among the multivariate versions of the Glivenko-Cantelli Lemma, Blum's version states the following:

For $\mu_n$ the empiric measure,

$$P^\infty \left\{ \lim_{n \to \infty} \sup_{A \in G} |\mu_n(A) - \mu(A)| = 0 \right\} = 1,$$

where $\mu$ must be absolutely continuous with respect to the corresponding Lebesgue measure in the Euclidean space $E$, and $G$ is some collection of half-spaces. In his paper, he also conjectures the validity of the theorem when dropping the requirement of absolute continuity of $\mu$. If his conjecture is correct, so is its automatic generalization by virtue of Theorem 3.2.

3.3 Strong Consistency of Quantile Estimators

As an example of immediate application of Theorem 3.2, the strong consistency of quantile estimators obtained from Rao-Blackwell distribution function estimators is shown in this section, when the parent is assumed to be absolutely continuous and one-to-one.
Definition

For \( \alpha \in (0, 1) \), the real number \( x_\alpha \) will be said to be the \( \alpha \)\textsuperscript{th} quantile of the distribution function \( F \) if,

\[
x_\alpha = \inf\{y; F(y) \geq \alpha\}.
\]

Let \( \tilde{F}_n(x) \) be a Rao-Blackwell distribution function estimator of \( F \), and define the quantile estimator \( \tilde{X}_\alpha \) to be the \( \alpha \)\textsuperscript{th} quantile of \( \tilde{F}_n(x) \) for each outcome of the sample.

First, it is shown that \( \tilde{X}_\alpha \) is not in general a Rao-Blackwell estimator, for which it suffices to show that:

\[
(\alpha+1)\tilde{X}_\alpha \neq E(\tilde{X}_\alpha \mid \theta^{n+1}),
\]

in general. For this, let

\[
F_\theta(x) = (x/\theta)I_{[0,\theta]}(x) + I(\theta, +\infty),
\]

(where \( I \) stands for an indicator function), where \( \theta \) is unknown.

\( \tilde{F}_n(x) \), obtained by conditioning on the minimal sufficient statistic \( X_{(n)} \) (the largest order statistic) is given by:

\[
\tilde{F}_n(x) = \begin{cases} 
(n - 1)/nX_{(n)}, & \text{for } 0 < x \leq X_{(n)}; \\
1, & \text{for } X_{(n)} < x.
\end{cases}
\]

Thus, the quantile estimator is given by:

\[
\tilde{X}_\alpha = \begin{cases} 
\alpha nX_{(n)}/(n-l), & \text{for } 0 < \alpha < (n-l)/n; \\
X_{(n)}, & \text{for } (n-l)/n \leq \alpha < l,
\end{cases}
\]

and
\[ E(\widetilde{X}_\alpha^\prime) = \begin{cases} \frac{an^2}{(n^2-1)}, & \text{for } 0 < \alpha < (n-1)/n; \\ \frac{n}{n+1}, & \text{for } (n-1)/n \leq \alpha. \end{cases} \]

It is clear then, that for any \( \alpha \in (0, 1) \),

\[ E(\widetilde{X}_\alpha^\prime) \neq E((n+k)\widetilde{X}_\alpha^\prime) \]

for some \( k \geq 1 \), i.e., the quantile estimators are not Rao-Blackwell estimators; therefore, for strong consistency of quantiles, Berk's result can not be used.

**Theorem 3.3**

Let \( \widetilde{F}_n^\prime \) be the Rao-Blackwell distribution function estimator of \( F \), obtained by conditioning on a symmetric sequential statistic. If \( F \) is absolutely continuous and one-to-one; then the quantile estimator \( \widetilde{X}_\alpha^\prime \) converges almost surely to the quantile \( x_\alpha \).

**Proof**

For any \( \varepsilon > 0 \),

\[ F(x_\alpha - \varepsilon) < F(x_\alpha) < F(x_\alpha + \varepsilon). \]

A necessary and sufficient condition for the almost sure convergence of \( \frac{\widetilde{X}_\alpha^\prime}{n} \) to \( x_\alpha \) is that for any \( \varepsilon > 0 \), there exists a set \( A_\varepsilon \) and an integer \( N_\varepsilon \) such that,

\[ P(A_\varepsilon) < \varepsilon \]

and for \( n \geq N_\varepsilon \).
\[ |\tilde{x}_n^{\alpha} - x_\alpha^{\alpha} | < \varepsilon \]

on \( A_{\varepsilon}^c \).

Fix \( \varepsilon > 0 \); \( F \) is one-to-one, so there exists

\[ \delta(\varepsilon) > 0 \text{ s.t. } F^{-1}(\alpha + 3\delta) - F^{-1}(\alpha - 3\delta) < \varepsilon. \]

For this \( \delta \), using the necessary and sufficient condition mentioned above, select the corresponding set for

\[ \delta^* = \min \{ \varepsilon, \delta \}, \]

\( A_{\delta^*} \), and an integer \( N_\delta \) for which:

\[ P(A_{\delta^*}) < \varepsilon, \]

and outside \( A_{\delta^*} \),

\[ \forall n \geq N_\delta \]

\[ |\tilde{F}_n(x) - F(x)| < \delta; \]

in particular,

\[ \tilde{F}_n(\tilde{x}_n^{\alpha}) - \delta < F(\tilde{x}_n^{\alpha}) < \tilde{F}_n(\tilde{x}_n^{\alpha}) + \delta, \forall n \geq N_\delta, \]

on \( A_{\delta^*}^c \). By its definition,

\[ \tilde{F}_n(\tilde{x}_n^{\alpha}) \geq \alpha, \]

thus,

\[ F(\tilde{x}_n^{\alpha}) > \alpha - \delta, \]
i.e.,

\[ \tilde{X}_{n}\alpha > \tilde{F}_{n}^{-1}(\alpha - \delta), \]

and on \( A_{\delta \gamma}^{c} \), the maximum jump that \( \tilde{F}_{n}(x) \) can have is clearly less than \( 2\delta \), for all \( n \geq N_\delta \).

Therefore,

\[ \tilde{F}_{n}(\tilde{X}_{n}\alpha - 0) > \tilde{F}_{n}(\tilde{X}_{\alpha}) - 2\delta, \]

hence,

\[ \alpha - 3\delta < F(\tilde{X}_{n}\alpha) < \alpha + 3\delta, \]

which yields:

\[ |\tilde{X}_{n}\alpha - x_\alpha| < \varepsilon, \ \forall \ n \geq N_\delta, \]

on \( A_{\delta \gamma}^{c} \). So let \( N_\varepsilon = N_\delta \) and \( A_{\varepsilon} = A_{\delta \gamma}^{c} \), which completes the proof.
4. GOODNESS-OF-FIT TESTS

4.1 Statement of the Problem

A goodness-of-fit problem arises when it is desired to test if a certain population produced a sample. The sample will be assumed to be independent. The population is identified with a probability measure, according to which the individual terms in the sample are distributed; or more generally, the population is identified with any probability measure belonging to some well defined class.

Definition

The following structure will be referred to as a goodness-of-fit problem:

"Given an independent sample $X_1, \ldots, X_n$ of observations on the random variable $X$ (scalar or vector valued), whose probability measure is $P$, defined on the Euclidean space $(\Omega, \mathcal{G})$, it is desired to test, $H_0 : P \in \mathcal{P}$, versus, $H_a : P \in \mathcal{D} - \mathcal{P}$, where $\mathcal{P}, \mathcal{D}$ are classes of probability measures on $(\Omega, \mathcal{G})$ and $\mathcal{P} \subset \mathcal{D}.$"

Usually, $\mathcal{P}$ will be a parametric class, and $\mathcal{D}$ will be unspecified. If strong assumptions can be made concerning the structure of $\mathcal{D}$, it may happen that the goodness-of-fit problem is solvable by means of Neyman-Pearson theory.

A goodness-of-fit problem can be characterized as simple or composite according to the class $\mathcal{P}$, i.e., if $\mathcal{P}$ is a singleton class, the problem will be said to be simple, otherwise it will be said to be composite.
Also, the dimensionality of the random observations will characterize a problem as univariate or multivariate.

Because of the regularity of conditional probabilities in Euclidean spaces, throughout this work all conditional distributions will be assumed to exist everywhere.

All problems considered will be understood to be univariate goodness-of-fit problems unless otherwise stated. Specifically, a multivariate problem is dealt with in Chapter 6.

In Section 2 of this chapter, a distribution-free test is proposed for the composite goodness-of-fit problem whose existence depends upon the particular class to be tested. Sufficient conditions are given on the class for the existence of such a test.

In Section 3, a different approach is followed by way of generalizing the probability integral transformation. The notion of distribution-freeness is analysed, and class-freeness is defined. Sufficient conditions are then given, on a class, for class-free, distribution-free statistics to exist. The usual absolute continuity requirement for a simple goodness-of-fit problem to admit a distribution-free statistic is reconstructed as a particular case of the sufficient conditions obtained in this section.

4.2 A Distribution-Free Test

For purposes of completeness, the definition of a distribution-free statistic for a composite goodness-of-fit problem is given.
Definition

A distribution-free statistic, for a given composite goodness-of-fit problem whose null class is $P$ is defined to be any measurable function of the observations, whose distribution is the same regardless of which $P \in P$ is true.

Note:

If the problem is simple, i.e., $P$ is a singleton class, then the above definition does not particularize to the distribution-freeness of a statistic for the corresponding problem.

Let $P$ be the null class of a composite goodness-of-fit problem and let $T_n$ (inducing $P^n$) be the minimal sufficient statistic for $P$. Denote $F_n$ the empirical distribution function and by $F_n^P$ the Rao-Blackwell distribution function estimator obtained by conditioning on $P^n$.

Definition

The generalized Kolmogorov-Smirnov statistic for the above problem is defined to be:

$$\sup_{x \in \mathbb{R}} \left| F_n^P(x) - F_n(x) \right| .$$

Srinivasan [21], has shown for the normal and exponential goodness-of-fit problems, that the above statistic is distribution-free. He also simulated its distribution for different values of $n$ for these cases.

Next, sufficient conditions on a composite problem for which the generalized Kolmogorov-Smirnov statistic is distribution-free are
derived. For this, let \( (\mathbb{R}^n, \mathcal{B}^n) \) be the \( n \)-dimensional product space of the sample \( X_1, \ldots, X_n \); as before, \( T_n \) (inducing \( \mathcal{B}^n \)) will denote the minimal sufficient statistic for \( \mathcal{P} \) and \( U_n \) (inducing \( \mathcal{C}^n \)) will denote the order statistics vector.

**Definition**

The class \( \mathcal{P} \) of measures on the Borel line \( (\mathbb{R}, \mathcal{B}) \) is said to admit an invariance transformation if there exists, for each \( \mathcal{P} \in \mathcal{P} \), a one to one function \( g_\mathcal{P} : \mathbb{R} \to \mathbb{R}_* \subset \mathbb{R} \), which induces a probability space \((\mathbb{R}_*, \mathcal{B}_*, \mathbb{P}_*)\) which is the same for all \( \mathcal{P} \in \mathcal{P} \).

The above definition can be relaxed, to allow \( g_\mathcal{P} \) to be essentially one to one, by which the following is meant:

Suppose \( g_\mathcal{P} \) is onto \( \mathbb{R}_* \), and denote by \( S_{y}^{-1} \) the inverse image of \( \{y\}, y \in \mathbb{R}_* \), i.e., \( S_{y}^{-1} = \{x \in \mathbb{R}; g_\mathcal{P}(x) = y\} \). Also, assume there exists a criteria; by means of which, for each \( y \in \mathbb{R}_* \), a "canonical" member of \( S_{y}^{-1} \) can be chosen. Denote this canonical member by \( y^{-1} \); then, if the set \( \bigcup_{y \in \mathbb{R}_*} (S_{y}^{-1} - \{y^{-1}\}) \) is measurable (\( \mathcal{B} \)) and has \( \mathbb{P} \) measure 0, \( g_\mathcal{P} \) will be said to be essentially one to one.

An example of an essentially one to one function \( g_\mathcal{P} \) is the probability integral transformation for \( \mathbb{P} \ll \lambda \), i.e., dominated by the Lebesgue measure on \((\mathbb{R}, \mathcal{B})\). In this case, the canonical element can be chosen as the infimum or supremum of the sets \( S_{y}^{-1} \).

In what follows, when a class \( \mathcal{P} \) is said to admit an invariance transformation, the functions \( g_\mathcal{P} \) will be assumed to be one to one. It can be shown that all results presented will also hold for essentially one to one functions.
Clearly, if \( \mathcal{P} \) admits an invariance transformation, then the class \( \mathcal{P}^{(n)} \) of product measures on \( (R^n, \mathcal{B}^n) \) also admits an invariance transformation, for if \( (g^*_P)_{P \in \mathcal{P}} \) is the set of one to one functions inducing \( (R^*_*, \mathcal{B}^*_*, P^*_*) \), then the functions: \( g^*_P: R^n \rightarrow R^n \), defined for each \( P^n \in \mathcal{P}^{(n)} \) by,

\[
  g^*_P(x_1, \ldots, x_n) = (g^*_P(x_1), \ldots, g^*_P(x_n))
\]

are one to one and induce the probability space \( (R^n, \mathcal{B}^n, P^n) \) which is the same for all \( P^n \in \mathcal{P}^{(n)} \).

Let \( \mathcal{G}^n, \mathcal{C}^n \) be the sub \( \sigma \)-algebras induced by the minimal sufficient statistic and the order statistics vector, respectively, for the family \( \mathcal{P} \).

**Condition 4.2.1**

Using the above notation, \( \mathcal{P} \) admits an invariance transformation and the direct mappings of \( \mathcal{C}^n, \mathcal{G}^n \) by means of \( g^*_P \) (which are themselves sub-\( \sigma \)-algebras of \( \mathcal{G}^*_n \)) are the same for all \( P \in \mathcal{P} \). Denote these \( \sigma \)-algebras by \( \mathcal{C}^n, \mathcal{G}^n \), respectively.

**Lemma 4.2.2**

Assume \( \mathcal{P} \) meets Condition 4.2.1, and let \( X \) be any quasi-integrable \( (\mathcal{P}) \) random variable in \( (R^n, \mathcal{G}^n) \). Then, for \( X^*_* \) obeying \( X = X^*_* g^*_P \),

\[
  E(X|\mathcal{G}^n) = E(X^*_*|\mathcal{G}^*_n) g^*_P \quad a.s.,
\]

\[
  E(X|\mathcal{C}^n) = E(X^*_*|\mathcal{C}^*_n) g^*_P \quad a.s.
\]
Proof

By definition, \( E(X|\mathcal{G}^n) \) is \( \mathcal{G}^n \) measurable; moreover,
\[
(\mathcal{G}_P^n)^{-1}(\mathcal{G}_x^n) = \mathcal{G}_x^n \text{; therefore, following Lemma 1, Lehmann [10], p. 37, there exists a } \mathcal{G}_x^n \text{ measurable real-valued function } h \text{ such that,}
\]
\[
(4.2.2.1) \quad E(X|\mathcal{G}^n) = h_{\mathcal{G}_P^n} \text{ a.s.}
\]

Using the definition of conditional expectations, and the fact that \( B \in \mathcal{G}^n \) iff \( B = (\mathcal{G}_P^n)^{-1}(B_*) \), \( B_* \in \mathcal{G}_x^n \), it follows that, \( \forall \, \mathcal{G}_x^n \in \mathcal{G}_x^n \), where \( B = (\mathcal{G}_P^n)^{-1}(B_*) \),
\[
(4.2.2.2) \quad \int_B E(X|\mathcal{G}^n) dP^n = \int_B X dP^n = \int_{B_*} X_* dP_* = \int_{B_*} E(X_*|\mathcal{G}_x^n) dP_*
\]
the second equality being an application of the transformation theorem.

From (4.2.2.1) and (4.2.2.2) it follows that,
\[
\forall \, B_* \in \mathcal{G}_x^n, \int_{B_*} h dP_* = \int_{B_*} E(X_*|\mathcal{G}_x^n) dP_*
\]
which by a.s. uniqueness of conditional expectations means that:
\[
E(X|\mathcal{G}^n) = E(X_*|\mathcal{G}_x^n) h_{\mathcal{G}_P^n} \text{ a.s.}
\]

Theorem 4.2.3

Consider the composite goodness-of-fit problem with null class \( \mathcal{P} \). Assume,

i) \( \mathcal{P} \) meets condition 4.2.1, and,

ii) \( \mathcal{G}_P \) is monotonic for each \( P \in \mathcal{P} \).
Then the statistic,
\[ \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - \hat{F}_n(x)|, \]
is distribution free; where,
\[ \hat{F}_n(x) = \mathbb{E}(\delta_x | \mathbb{G}_n^n), \hat{F}_n(x) = \mathbb{E}(\delta_x | \mathbb{C}_n^n). \]

**Proof**

Fix \( P \in \mathcal{P} \) arbitrarily, and consider the transformed probability space \((\mathbb{R}_x^n, \mathbb{G}_x^n, \mathbb{P}_x^n)\). In this space, due to the monotonicity of \( g_P \), for \( X = \delta_x \) in the original space, \( X^* \) is given by \( \delta^*_y \), where:
\[
\delta^*_y = \begin{cases} 
\delta_{g_P(x)}, & \text{if } g_P \text{ is increasing;} \\
1 - \delta_{g_P(x)}, & \text{if } g_P \text{ is decreasing}; 
\end{cases}
\]
and,
\[
\delta_x = \delta^*_y \log_P. 
\]

Applying Lemma 4.2.2,
\[ \hat{F}_n(x) = \mathbb{E}(\delta^*_y | \mathbb{G}_x^n) \log_P \mathbb{G}_P^n \text{ a.s.}, \quad \text{and} \]
\[ \hat{F}_n(x) = \mathbb{E}(\delta^*_y | \mathbb{G}_x^n) \log_P \mathbb{G}_P^n \text{ a.s.} \]

Therefore, the r.v. \(|\hat{F}_n(x) - \hat{F}_n(x)|\) can be written as,
\[ |\mathbb{E}(\delta^*_y | \mathbb{G}_x^n) \log_P \mathbb{G}_P^n - \mathbb{E}(\delta^*_y | \mathbb{C}_x^n) \log_P \mathbb{C}_P^n|, \]
and thus the probability in the original space,
\[ \mathbb{P}_{\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - \hat{F}_n(x)| \leq k} \]
is equal to the probability

\[ P^\bullet \left[ \sup_{x \in \mathbb{R}} \left| E(\delta^*_y | x^*_n) - E(\delta^*_y | c^*_n) \right| \leq k \right], \]

and clearly this last statement is equivalent to,

\[ P^\bullet \left[ \sup_{y \in \mathbb{R}} \left| E(\delta^*_y | x^*_n) - E(\delta^*_y | c^*_n) \right| \leq k \right], \]

since the supremum can be taken over the images \( g_\mathcal{P}(x) \).

The two cases given by Srinivasan [21], viz., the normal and exponential case are immediate by application of the above result. The functions \( g_\mathcal{P} \) used in his cases are the "standarizing" functions. In his paper, however, there is a mistake in the claim that,

\[ \sup_{x \in \mathbb{R}} | \tilde{F}_n(x) - F_n(x) | \]

is the same as

\[ \max_{i=1, \ldots, n} | \tilde{F}_n(x_i) - F_n(x_i) |, \]

where \( x_1, \ldots, x_n \) is the observed sample; since the supremum in the first expression need not be attained.

**Note on Consistency**

From Theorem 3.2, the generalized Kolmogorov-Smirnov statistic converges to 0 a.s., whenever the null hypothesis is true.

The question of sensitivity or limiting behaviour under the alternative hypothesis (meaning any \( F^1 \notin \mathcal{P} \)) is in general a difficult one. It can be asserted, however, that if the minimal sufficient statistic for \( \mathcal{P} \) is not also sufficient for \( \mathcal{P} + \{ F^1 \} \), then \( \tilde{F}_n(x) \) can
not be expected to be an adequate estimate of \( F_{\infty}(x) \). Thus, under such alternatives, the generalized Kolmogorov-Smirnov statistic will be expected to be "larger".

With respect to limiting behaviour of \( \hat{F}_n(x) \) under an alternative \( F^1 \), no easy answer seems to be available. This problem, which in essence is the limiting behaviour of a conditional expectation when the probability measure was wrongly assumed requires some attention.

4.3 The Conditional Probability Integral Transformation

Consider a simple goodness-of-fit problem whose null class corresponds to the distribution function \( F_0 \).

As remarked in Section 2, distribution-freeness of a statistic for a simple problem is not a particular case of that for a composite problem. Following the definition given by Birnbaum [2], in the simple problem a statistic is said to be distribution-free if the distribution of the test statistic is the same for different null classes.

Let \( \mathcal{H} \) be the collection of all singleton null classes whose corresponding distribution functions are absolutely continuous. A well known result is that the Kolmogorov-Smirnov statistic,

\[
\sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|,
\]

is distribution-free for all absolutely continuous \( F_0 \) (i.e., with respect to \( \mathcal{H} \)).

The above assertion can be shown very easily using the probability integral transformation. For any absolutely continuous
parent \( F_0 \), the random variables \( F_0(x_1), \ldots, F_0(x_n) \) are independently and uniformly distributed in the \( n \)-dimensional unit hypercube.

Obviously, such a transformation leaves invariant the \( \sigma \)-algebra induced by the order statistics, and thus, \( F_0 \) being monotonic, by Lemma 4.2.2,

\[
F_n(x) = F_n^*(y)_{\circ F_0^n},
\]

where \( F_0^n : \mathbb{R}^n \to (0, 1)^n \) is given by \( (F_0(x_1), \ldots, F_0(x_n)) \) and \( F_n^*(y) \) is the empirical distribution function computed from the transformed independent "sample" \( F_0(x_1), \ldots, F_0(x_n) \) and evaluated at \( y = F_0(x) \). Therefore, by an argument similar to that used in theorem 4.2.3,

\[
\sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|
\]

is identically distributed with,

\[
\sup_{y \in (0, 1)} |F_n^*(y) - y|,
\]

and this is true for all absolutely continuous \( F_0 \).

In the remainder of this section, the composite problem will be studied. The main idea will be that of transforming the independent observations into \( U(0, 1) \) independent random variables. Unlike the simple problem, the relationship between the statistic computed in the transformed space and that computed in the beginning space will not be studied, except for a chi-square type statistic proposed in Chapter 5.

Let \( m \) be a collection of null classes \( \mathcal{P} \), each of which identifies a composite goodness-of-fit problem. First, the class-freeness property of a statistic is introduced.
Definition

A class-free \((\mathfrak{m})\) statistic for solving the composite goodness-of-fit problem whose null class is \(\mathcal{P} \in \mathfrak{m}\) is a distribution-free \((\mathcal{P})\) statistic whose unique distribution is invariant under choices of classes \(\mathcal{P} \in \mathfrak{m}\).

First, note that the definition of a distribution-free statistic for the simple problem is a particular case of the above definition, namely a class-free \((\mathfrak{m})\) statistic, where \(\mathfrak{m}\) is the set of all singleton classes \(\{\mathcal{P}_0\}\). (Usually where \(\mathcal{P}_0 \ll \lambda\) in the univariate case.)

Let \(\mathcal{P}\) be a null class. Consider the minimal sufficient statistic for \(\mathcal{P}\) based on \(n\) independent observations. Denote by \(\mathfrak{G}_n\) the \(\sigma\)-algebra induced by this statistic and let \(\hat{F}_n(x)\) be the corresponding Rao-Blackwell estimating distribution function. Denote by \(\mathfrak{F}\) the class of distribution functions corresponding to \(\mathcal{P}\). A well known result is that:

Result 4.3.1

If \(F \in \mathfrak{F}\) is the parent and \(F\) is not absolutely continuous, then with positive probability \(\hat{F}_n(x)\) will not be absolutely continuous.

Proof

Denote by \(P\) the probability measure associated with \(F\) and by \(\tilde{P}_n\) the conditional probability measure associated with \(\tilde{F}_n\). Since \(P \ll \lambda\), it means that there exists a Borel set \(D\) for which \(\lambda(D) = 0\), and \(P(D) > 0\). Thus, since
\[ E(\widehat{F}_n(D)) = P(D) > 0, \]
\[ \Rightarrow \widehat{F}_n(D) > 0, \]

with positive probability.

From the above result it can be asserted that if \( P \in \mathcal{F} \), is the true parent, and if \( \widehat{F}_n \) is absolutely continuous a.s., then, \( P \ll \lambda \).

**Theorem 4.3.2**

Let \( X_1, \ldots, X_n \) be an independent sample from the univariate parent \( P \in \mathcal{F} \). Let \( T_n \) be any sufficient statistic for \( \mathcal{F} \) and \( \widehat{F}_n \) the corresponding Rao-Blackwell distribution estimator of \( F \). If \( \widehat{F}_n \) is absolutely continuous a.s., then,

\[ \widehat{F}_n(X_i) \sim U(0, 1), \quad \text{for } i = 1, \ldots, n. \]

**Proof**

For fixed \( T_n \), \( \widehat{F}_n \) is the conditional distribution function of \( X_i \); thus by means of the probability integral transformation the conditional distribution of \( \widehat{F}_n(X_i) \) given \( T_n \) is uniform in \((0, 1)\) a.s.; i.e.,

\[ \widehat{F}_n(X_i) \sim U(0, 1). \]

Note that \( \widehat{F}_n(X_1), \ldots, \widehat{F}_n(X_n) \) are not necessarily independent; in fact, for given \( T_n \), the conditional distribution of \( X_1, \ldots, X_n \) will, in general, be singular.
In order to overcome this difficulty, a result from Rosenblatt [17], will be used. In that paper it is shown that for a k-variate absolutely continuous distribution function \( F(x_1, \ldots, x_k) \), the k transformations:

\[
P[X_1 \leq x_1] = F(x_1),
\]

\[
P[X_2 \leq x_2 | X_1 = x_1] = F(x_2 | x_1), \text{ etc., and,}
\]

\[
P[X_k \leq x_k | X_1 = x_1, \ldots, X_{k-1} \leq x_{k-1}] = F(x_k | x_1, \ldots, x_{k-1}),
\]

are such that the k r.v.'s,

\[
F(X_1), F(X_2 | X_1), \ldots, F(X_k | X_1, \ldots, X_{k-1})
\]

are uniformly and independently distributed in \((0, 1)^k\).

With the above multivariate probability integral transformation and the conditional probability integral transformation, it will be shown that under some conditions of absolute continuity one can obtain a \( U(0, 1) \) independent sample, but first, a characterization is given for a random multivariate transformation.

Let \( \widetilde{F}(x_1, x_2) \) be the conditional bivariate joint distribution of \( X_1, X_2 \) given some statistic \( T \), i.e.,

\[
\widetilde{F}(x_1, x_2) = E[I_{[X_1 \leq x_1, X_2 \leq x_2]} | T).
\]

For each value of \( T = t \), assume \( \widetilde{F}_t(x_1, x_2) \) is absolutely continuous, and let \( \widetilde{F}_t(x_2 | x_1) \) be the conditional distribution of \( X_2 \) obtained from \( \widetilde{F}_t(x_1, x_2) \) for given \( X_1 = x_1 \). For unspecified \( t \), denote this random variable by \( \widetilde{F}(x_2 | x_1) \).
Lemma 4.3.3

\( \tilde{F}(x_2| x_1) \) coincides a.s. with the conditional distribution of \( X_2 \) given \( X_1 \) and \( T \), evaluated at \( X_1 = x_1 \), i.e.,

\[
\tilde{F}(x_2| x_1) = E(I_{[X_2 \leq x_2]}| T, X_1 = x_1) \quad \text{a.s.}
\]

Proof

Denote by \( \tilde{F}(x_1) \) the marginal d.f. of \( \tilde{F}(x_1, x_2) \) computed for each \( T = t \), i.e.,

\[
\tilde{F}(x_1) = \lim_{x_2 \to \infty} \tilde{F}(x_1, x_2) = \lim_{x_2 \to \infty} E(I_{[X_1 \leq x_1, X_2 \leq x_2]}| T).
\]

By the monotone convergence theorem for conditional expectations the last term in the above expression is equal to

\[
E(I_{[X_1 \leq x_1]}| T).
\]

Therefore,

\[
\tilde{F}(x_1) = E(I_{[X_1 \leq x_1]}| T).
\]

From its definition, for each value of \( T \), \( \tilde{F}(x_2| x_1) \) obeys the following relationship:

\[
(4.3.3.1) \quad \tilde{F}(x_1, x_2) = \int_{-\infty}^{x_1} \tilde{F}(u) d\tilde{F}(u).
\]

Now, \( \tilde{F}(x_1, x_2) \) being a conditional expectation, must obey, for every \( B_T \in \sigma(T) \), the following relationship:

\[
(4.3.3.2) \quad \int_{B_T} \tilde{F}(x_1, x_2) dP_T = \int_{B_T} I_{[X_1 \leq x_1, X_2 \leq x_2]} dP = P([X_1 \leq x_1, X_2 \leq x_2]_{B_T}).
\]
Also, \( E(I_{X_2 \leq x_2} | T, X_1) \) obeys, for every \( B_T \in \sigma(T) \),

\[
\int_{B_T} \int_{-\infty}^{x_1} E(I_{X_2 \leq x_2} | T, X_1) dP_{T, X_1} = P([X_1 \leq x_1, X_2 \leq x_2] \cap B_T).
\]

From (4.3.3.1), (4.3.3.2), (4.3.3.3), and the fact that

\[
dP_{T, X_1} = dP_{X_1}^T dP_T
\]

where obviously,

\[
P_{X_1}^T (-\infty, u) = \tilde{F}(u);
\]

for every \( B_T \in \sigma(T) \), and any \( x_1 \in \mathbb{R} \),

\[
\int_{B_T} \int_{-\infty}^{x_1} \tilde{F}(x_2 | u) dP_{X_1}(u) dP_T = \int_{B_T} \int_{-\infty}^{x_1} E(I_{X_2 \leq x_2} | T, X_1 = u) dP_{X_1}(u) dP_T,
\]

therefore

\[
\tilde{F}(x_2 | x_1) = E(I_{X_2 \leq x_2} | T, X_1 = x_1) \text{ a.s.}
\]

With the above characterization, the following theorem is presented without a detailed description of the marginal and conditional distribution of a joint distribution which itself will be a Rao-Blackwell conditional distribution. Clearly, Lemma 4.3.3 generalizes to the k-variate case.

**Theorem 4.3.4**

Let \( X_1, \ldots, X_n \) be an independent sample from \( P \in \mathcal{P} \). Let \( \tilde{F}(x_1, \ldots, x_n), \ell \leq n \) be the joint conditional distribution of the first \( \ell \) terms of the sample given some sufficient statistic \( T_n = T_n(X_1, \ldots, X_n) \).
If $\tilde{F}_n(x_1, \ldots, x_\ell)$ is absolutely continuous a.s. $[\mathcal{R}]$, then the r.v.'s,

$$\tilde{F}_n(x_1), \tilde{F}_n(x_2|x_1), \ldots, \tilde{F}_n(x_\ell|x_1, \ldots, x_{\ell-1})$$

are independent and uniformly distributed in $(0, 1)\ell$.

Proof

For each value of $T_n$ (a.s.), the joint conditional distribution of $X_1, \ldots, X_\ell$ is absolutely continuous, therefore, by the multivariate probability integral transformation, the $\ell$ above proposed r.v.'s are conditionally distributed uniformly in the unit $\ell$-dimensional hypercube.

Since its conditional distribution does not depend on the particular value of $T_n$ a.s., then the assertion of the theorem follows.

The absolute continuity (a.s. $[\mathcal{R}]$) requirement of the joint conditional distribution used in the above theorem, as well as the assumption (implicit) of $\ell > 0$ motivates the following definition:

Definition

Let $\mathcal{P}$ be a family of univariate probability measures and $X_1, \ldots, X_n$ an independent sample from some $P \in \mathcal{P}$. Let $T_n$ be a sufficient statistic for $\mathcal{P}$ based on $X_1, \ldots, X_n$. Denote by $\tilde{F}_n(x_1, \ldots, x_\ell)$ the joint conditional distribution function of the first $\ell$ terms of the sample given $T_n$. 
The maximum \( \ell (= 0, 1, \ldots, n) \) for which \( \tilde{F}_n(x_1, \ldots, x_\ell) \) is absolutely continuous a.s. \( [\mathcal{P}] \) will be called the **absolute continuity rank of \( \mathcal{P} \) given \( T_n \). The arbitrariness of selecting the "first \( \ell \) observations" disappears when \( T_n \) is required to be symmetric (which is always the case for independent identically distributed elements in the sample).

Note that if \( \mathcal{P} \) is not a subset of the class of all absolutely continuous probability measures with respect to Lebesgue measure, then the absolute continuity rank given any sufficient statistic must be 0, by virtue of Result 4.3.1.

The question of how large the absolute continuity rank of a class \( \mathcal{P} \) can get, for alternative sufficient statistics is answered with the following argument.

Seheult and Quesenberry [19] have shown that a necessary and sufficient condition for \( \tilde{F}_n(x_1, \ldots, x_\ell) \) to be absolutely continuous, is that there exists a \( \sigma(T_n) \) measurable unbiased estimator of the density of \( X_1, \ldots, X_\ell \). Note that if the absolute continuity rank of \( \mathcal{P} \) given \( T_n \) is \( \ell > 0 \), then the existing \( \sigma(T_n) \) measurable unbiased estimator can be conditioned with \( Z_n \), the minimal sufficient statistic for \( \mathcal{P} \), thus yielding a \( \sigma(Z_n) \) measurable unbiased density estimator. The absolute continuity rank of \( \mathcal{P} \) given any sufficient statistic is, therefore, less than or equal to that given the minimal sufficient statistic.

In view of this result, when considering a class \( \mathcal{P} \), the use of the minimal sufficient statistic will be understood in computing the Rao-Blackwell estimators.
Let \( \mathcal{M} \) be the collection of classes \( \mathcal{P} \) for which the absolute continuity rank (given the corresponding minimal sufficient statistic) is \( \lambda (> 0) \). Because of Theorem 4.3.4, any measurable function of the \( \lambda \) transformed observations will be a class-free (\( \mathcal{M} \)) statistic. Theorem 4.3.4, in essence, reduces the composite problem to a simple problem.

For any problem whose null class is in \( \mathcal{M} \), it is true that under the null hypothesis the \( \lambda \) transformed observations will be uniformly distributed on \((0, 1)\). The question arises as to the distribution of these observations under alternatives, and the answer is the same as that given at the end of Section 2 of this chapter. It is to be expected that if the transformations used are "far" from the correct ones, then the observations will not be uniformly distributed on \((0, 1)\).

In Chapter 3, the notion of sequential statistics was introduced; here, a slightly stronger notion is given.

**Definition**

For each \( n \geq 1 \), let \( T_n(X_1, \ldots, X_n) \) be a statistic based on the independent sample \( X_1, \ldots, X_n \); \( (T_n)_{n \geq 1} \) will be said to be **doubly sequential**, if,

\[
\sigma(T_n', X_n) = \sigma(T_{n-1}', X_n).
\]

Next, an interesting result is presented by means of which the transformations proposed in Theorem 4.3.4 take a very simple form, and for which convergence of the class-free statistic depends entirely
upon the choice of the statistic and not on the random transformations.

Theorem 4.3.5

Let $X_1, \ldots, X_n$ be an independent sample according to $P \in \mathcal{P}$, the minimal sufficient statistic $T_n$ for $P$ be doubly sequential, and $\ell (> 0)$ be the absolute continuity rank for $P$ (given $T_n$). Then the r.v.'s:

$$
\tilde{F}_{n-\ell+1}(X_{n-\ell+1}), \ldots, \tilde{F}_{n-1}(X_{n-1}),
$$

$\tilde{F}_n(X_n)$ are uniformly distributed on the unit hypercube $(0, 1)^\ell$, where,

$$
\tilde{F}_j(x_j) = E(I_{X_j \leq x_j} | T_j) \text{ for } j = n - \ell + 1, \ldots, n.
$$

Proof

Select the transformations of Theorem 4.3.4, in reverse order; a set of $\ell$ random variables is obtained which are uniformly distributed in $(0, 1)^\ell$.

Thus it suffices to show that the $j^{th}$ variable, obtained by the transformation, $\tilde{F}_n(x_j | X_n, \ldots, x_{j+1})$, corresponds to $\tilde{F}_j(X_j)$.

From Lemma 4.3.3,

$$
\tilde{F}_n(x_j | X_n, \ldots, X_{j+1}) = E(I_{X_j \leq x_j} | T_n, X_n, \ldots, X_{j+1}) \text{ a.s.}
$$

Now $T_n$ being doubly sequential implies that the right hand side of the above equality is a.s. equal to $\tilde{F}_j(x_j)$. 
Corollary 1

Under the assumptions of Theorem 4.3.5, let \( L = n - v \) (where \( v \) is constant) and suppose that \( Y_{v+1}', \ldots, Y_n' \) are the transformed observations corresponding to a sample of size \( n \). If a sample of size \( (n + 1) \) is considered by adding an independent observation to the original sample, then the corresponding transformed observations are \( Y_{v+1}', \ldots, Y_n', Y_{n+1}' \).

Because of the above result, even though there seems to be no reason for setting the order of the transformations in any specific way, under double sequentiality of the minimal sufficient statistic, and under reverse order selection, the convergence properties of the test statistic used on the transformed observations turns out to be clearly independent of the convergence properties of the transformations.

It should be clear at this stage that multivariate goodness-of-fit problems can be solved also by transforming a subset of the sample into \( U(0, 1) \) random variables. To show this, let \( \mathcal{P} \) be a family of \( k \)-variate probability measures.

For \( X_1, \ldots, X_n \) an independent sample of vectors distributed according to \( P \in \mathcal{P} \), assume that the absolute continuity rank of \( P \) (w.r. to \( \lambda^k \)) given the minimal sufficient statistic is \( L(> 0) \). For simplicity, assume the minimal sufficient statistic to be doubly sequential, also.
Corollary 2

Under the above assumptions, there exist \( k \cdot \ell \) random transformations which map the random matrix \([X_{n-\ell+1}, \ldots, X_n]\) into \( k \cdot \ell \) uniformly independently and identically distributed r.v.'s on \((0, 1)\).

Proof

Consider \( \tilde{F}_n (x_{n-\ell+1}, \ldots, x_n) \), and note that for each \( j = n - \ell + 1, \ldots, n; \)

\[
\tilde{F}_n(x_j | x_n, \ldots, x_{j+1}) = \tilde{F}_j(x_j) \text{ a.s.,}
\]

then consider the \( k \) transformations that are obtained from \( \tilde{F}_j(x_j) \), for each of the \( \ell \) values of \( j \).
5. CHI-SQUARE TYPE STATISTICS

In this chapter, some results concerning the chi-square goodness-of-fit statistic are presented.

Lancaster, [9] p. 3 says:

"Pearson's contributions to statistical theory were numerous, but perhaps the greatest of them was the \( \chi^2 \) test of goodness-of-fit ...".

This chapter is an application of the theory developed in Chapter 4, for the case when the statistic chosen for the transformed observations is a chi-square statistic. It will be shown that in this important case, the test can be carried out in the original space, with random selection of cells.

Since exact distribution theory for finite samples is the main appeal of the transformations developed in Chapter 4, a review of recent work in the area of simple goodness-of-fit problems is given in the next section.

5.1 Simple Goodness-of-Fit Problem

Consider a simple goodness-of-fit problem, whose null hypothesis is identified by the absolutely continuous distribution function \( F_o \). Denote by \( F_n \) the empirical distribution function. The chi-square statistic (\( \chi^2 \)), for testing \( F_o \), with \( k \) cells defined by the partition:

\[ \{(c_{i-1}, c_i]\}_{i=1, \ldots, k}, \quad \text{with} \quad -\infty = c_0 < c_1 < \cdots < c_k = +\infty ; \]

is defined to be:
\[ \chi^2 = \sum_{i=1}^{k} \frac{n}{F_0(C_i) - F_0(C_{i-1})} \left( F_n(C_i) - F_n(C_{i-1}) - F_0(C_i) + F_0(C_{i-1}) \right)^2. \]

For a given \( k \) and a given partition, the asymptotic distribution of \( \chi^2 \) is that of \( \chi^2_{(k-1)} \). Different values of \( k \) and different selections of the partition yield, in general, different finite sample properties of the test.

The arbitrariness of \( k \) and \( \{C_i\}_{i=1}^{k-1} \) has induced considerable research. A wide variety of recommendations appear in the literature regarding the choice of cells and the number of them. These recommendations were developed mainly to achieve a good approximation to the limiting distribution.

Mann and Wald, [14], introduced the idea of using equiprobable cells (under \( H_0 \)) and more recently, Good, et al [8] has mentioned that the selection of cells proposed by Mann and Wald yields a distribution-free statistic. These writers also computed the exact distribution of \( \chi^2 \) for this case.

It is easy to see that equiprobable cells yield a distribution-free statistic, since under these circumstances, the statistic \( \chi^2 \) is identical to the random variable \( T_{n,k} \) given by:

\[ T_{n,k} = \frac{k}{n} \sum_{i=1}^{k} N_i^2 - n, \]

where \( (N_1, \ldots, N_k) \) have a joint multinomial distribution with

\[ p_i = \frac{1}{k}, \quad i = 1, \ldots, k \quad \text{and} \quad \sum_{i=1}^{k} N_i = n. \]
Table 5.1  Table of critical points of the distribution of $T_n$ for the closest above and below significance levels to $\cdot\cdot05$

<table>
<thead>
<tr>
<th>n</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.000(.333)</td>
<td>2.000(.250)</td>
<td>3.000(.200)</td>
<td>4.000(.167)</td>
</tr>
<tr>
<td></td>
<td>4.000(0.000)</td>
<td>6.000(0.000)</td>
<td>8.000(0.000)</td>
<td>10.000(0.000)</td>
</tr>
<tr>
<td>3</td>
<td>2.000(0.111)</td>
<td>3.667(0.063)</td>
<td>5.333(0.040)</td>
<td>3.000(0.444)</td>
</tr>
<tr>
<td></td>
<td>6.000(0.000)</td>
<td>9.000(0.000)</td>
<td>12.000(0.000)</td>
<td>7.000(0.028)</td>
</tr>
<tr>
<td>4</td>
<td>2.000(0.333)</td>
<td>4.000(0.203)</td>
<td>6.000(0.136)</td>
<td>8.000(0.097)</td>
</tr>
<tr>
<td></td>
<td>3.500(0.037)</td>
<td>6.000(0.016)</td>
<td>8.500(0.008)</td>
<td>11.000(0.005)</td>
</tr>
<tr>
<td>5</td>
<td>2.800(0.136)</td>
<td>5.400(0.063)</td>
<td>6.000(0.098)</td>
<td>8.200(0.059)</td>
</tr>
<tr>
<td></td>
<td>5.200(0.012)</td>
<td>8.600(0.004)</td>
<td>8.000(0.034)</td>
<td>10.600(0.020)</td>
</tr>
<tr>
<td>6</td>
<td>4.000(0.053)</td>
<td>6.000(0.063)</td>
<td>5.667(0.098)</td>
<td>8.000(0.059)</td>
</tr>
<tr>
<td></td>
<td>7.000(0.034)</td>
<td>7.333(0.019)</td>
<td>9.000(0.028)</td>
<td>12.000(0.014)</td>
</tr>
<tr>
<td>8</td>
<td>4.750(0.059)</td>
<td>6.000(0.095)</td>
<td>8.250(0.054)</td>
<td>8.500(0.063)</td>
</tr>
<tr>
<td></td>
<td>6.250(0.033)</td>
<td>7.000(0.034)</td>
<td>9.500(0.037)</td>
<td>10.000(0.048)</td>
</tr>
<tr>
<td>10</td>
<td>5.600(0.059)</td>
<td>6.800(0.052)</td>
<td>8.000(0.055)</td>
<td>9.200(0.076)</td>
</tr>
<tr>
<td></td>
<td>6.200(0.022)</td>
<td>7.600(0.037)</td>
<td>9.000(0.040)</td>
<td>10.400(0.040)</td>
</tr>
<tr>
<td>12</td>
<td>4.500(0.070)</td>
<td>6.667(0.065)</td>
<td>8.000(0.074)</td>
<td>10.000(0.060)</td>
</tr>
<tr>
<td></td>
<td>6.000(0.048)</td>
<td>7.333(0.048)</td>
<td>8.833(0.040)</td>
<td>11.000(0.035)</td>
</tr>
<tr>
<td>14</td>
<td>5.286(0.068)</td>
<td>7.143(0.060)</td>
<td>8.857(0.052)</td>
<td>10.000(0.059)</td>
</tr>
<tr>
<td></td>
<td>6.143(0.033)</td>
<td>7.714(0.046)</td>
<td>9.571(0.034)</td>
<td>10.857(0.038)</td>
</tr>
<tr>
<td>16</td>
<td>5.375(0.061)</td>
<td>6.500(0.061)</td>
<td>9.000(0.052)</td>
<td>10.250(0.054)</td>
</tr>
<tr>
<td></td>
<td>6.125(0.039)</td>
<td>7.500(0.043)</td>
<td>9.625(0.035)</td>
<td>11.000(0.036)</td>
</tr>
<tr>
<td>18</td>
<td>5.333(0.057)</td>
<td>7.333(0.051)</td>
<td>8.667(0.060)</td>
<td>10.000(0.063)</td>
</tr>
<tr>
<td></td>
<td>6.333(0.033)</td>
<td>7.778(0.039)</td>
<td>9.222(0.043)</td>
<td>10.667(0.046)</td>
</tr>
<tr>
<td>20</td>
<td>5.200(0.056)</td>
<td>7.200(0.060)</td>
<td>9.000(0.051)</td>
<td>10.000(0.062)</td>
</tr>
<tr>
<td></td>
<td>6.100(0.045)</td>
<td>7.600(0.042)</td>
<td>9.500(0.039)</td>
<td>10.600(0.048)</td>
</tr>
<tr>
<td>$\infty$</td>
<td>5.991(0.050)</td>
<td>7.815(0.050)</td>
<td>9.488(0.050)</td>
<td>11.070(0.050)</td>
</tr>
</tbody>
</table>
Table 5.1  (Continued)

<table>
<thead>
<tr>
<th>n</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5.000 (.143)</td>
<td>6.000 (.125)</td>
<td>7.000 (.111)</td>
<td>8.000 (.100)</td>
</tr>
<tr>
<td></td>
<td>12.000 (.000)</td>
<td>14.000 (.000)</td>
<td>16.000 (.000)</td>
<td>18.000 (.000)</td>
</tr>
<tr>
<td>3</td>
<td>4.000 (.368)</td>
<td>5.000 (.344)</td>
<td>6.000 (.304)</td>
<td>7.000 (.280)</td>
</tr>
<tr>
<td></td>
<td>8.666 (.020)</td>
<td>10.333 (.016)</td>
<td>12.000 (.012)</td>
<td>13.667 (.010)</td>
</tr>
<tr>
<td>4</td>
<td>10.000 (.073)</td>
<td>12.000 (.057)</td>
<td>9.500 (.078)</td>
<td>11.000 (.064)</td>
</tr>
<tr>
<td></td>
<td>13.500 (.003)</td>
<td>16.000 (.002)</td>
<td>14.000 (.045)</td>
<td>16.000 (.037)</td>
</tr>
<tr>
<td>5</td>
<td>7.600 (.163)</td>
<td>9.400 (.128)</td>
<td>11.200 (.104)</td>
<td>18.000 (.086)</td>
</tr>
<tr>
<td></td>
<td>10.400 (.038)</td>
<td>12.600 (.026)</td>
<td>14.800 (.018)</td>
<td>17.000 (.014)</td>
</tr>
<tr>
<td>6</td>
<td>8.000 (.145)</td>
<td>10.000 (.103)</td>
<td>12.000 (.075)</td>
<td>14.000 (.057)</td>
</tr>
<tr>
<td></td>
<td>10.333 (.038)</td>
<td>12.666 (.026)</td>
<td>15.000 (.018)</td>
<td>17.333 (.014)</td>
</tr>
<tr>
<td>8</td>
<td>9.500 (.105)</td>
<td>12.000 (.067)</td>
<td>12.250 (.089)</td>
<td>14.500 (.066)</td>
</tr>
<tr>
<td></td>
<td>11.250 (.034)</td>
<td>14.000 (.020)</td>
<td>14.500 (.045)</td>
<td>17.000 (.031)</td>
</tr>
<tr>
<td>10</td>
<td>11.000 (.059)</td>
<td>12.400 (.059)</td>
<td>13.400 (.060)</td>
<td>14.000 (.071)</td>
</tr>
<tr>
<td></td>
<td>12.400 (.036)</td>
<td>14.000 (.035)</td>
<td>15.200 (.037)</td>
<td>16.000 (.039)</td>
</tr>
<tr>
<td>12</td>
<td>10.200 (.083)</td>
<td>12.000 (.072)</td>
<td>13.500 (.069)</td>
<td>14.667 (.066)</td>
</tr>
<tr>
<td></td>
<td>11.333 (.049)</td>
<td>13.333 (.042)</td>
<td>15.000 (.040)</td>
<td>16.333 (.043)</td>
</tr>
<tr>
<td>14</td>
<td>11.000 (.068)</td>
<td>12.286 (.070)</td>
<td>14.286 (.051)</td>
<td>14.571 (.073)</td>
</tr>
<tr>
<td></td>
<td>12.000 (.043)</td>
<td>13.429 (.045)</td>
<td>15.571 (.088)</td>
<td>16.000 (.046)</td>
</tr>
<tr>
<td>16</td>
<td>11.125 (.065)</td>
<td>13.000 (.056)</td>
<td>14.375 (.056)</td>
<td>15.250 (.064)</td>
</tr>
<tr>
<td></td>
<td>12.000 (.048)</td>
<td>14.000 (.038)</td>
<td>15.500 (.035)</td>
<td>16.500 (.043)</td>
</tr>
<tr>
<td>18</td>
<td>11.556 (.054)</td>
<td>13.111 (.056)</td>
<td>14.000 (.061)</td>
<td>15.333 (.061)</td>
</tr>
<tr>
<td></td>
<td>12.333 (.043)</td>
<td>14.000 (.040)</td>
<td>15.000 (.046)</td>
<td>16.444 (.045)</td>
</tr>
<tr>
<td>20</td>
<td>11.500 (.061)</td>
<td>12.800 (.062)</td>
<td>14.200 (.061)</td>
<td>16.000 (.052)</td>
</tr>
<tr>
<td></td>
<td>12.200 (.046)</td>
<td>13.600 (.045)</td>
<td>15.100 (.045)</td>
<td>17.000 (.039)</td>
</tr>
<tr>
<td>∞</td>
<td>12.592 (.050)</td>
<td>14.067 (.050)</td>
<td>15.507 (.050)</td>
<td>16.919 (.050)</td>
</tr>
</tbody>
</table>
In that the distributions of $T_{n,k}$ will be used in Section 2; these were obtained for some values of \( n \) and \( k \), and a table of .05 critical points is exhibited.

These distributions are discrete, thus the .05 level is not necessarily achieved, therefore, for the closest above and below significance levels the critical points appear together with the corresponding critical level.

The tabulations given by Good, et al \[8\] are actually those of a quadratic approximation to the $T_{n,k}$ distribution. Zahn and Roberts \[26\] give the exact $T_{n,k}$ distribution for \( n=k \) (cell expectations of one) for \( n = 1(1)25(5)40,50 \).

5.2 Composite Goodness-of-Fit Problem

Theorem 5.2

For the composite goodness-of-fit problem with null class \( \Phi \), let the absolute continuity rank of \( \Phi \) given the minimal sufficient statistic be \( \ell (> 0) \), and let \( p_1, \ldots, p_k \) be fixed positive probabilities, \( \sum p_j = 1 \). Then there exists \( \ell \) random partitions of \( \mathbb{R} \), given by:

\[
\{(\tilde{C}_{i-1,j}, \tilde{C}_{i,j}) \}_{i=1}^k,
\]

where

\[-\infty = \tilde{C}_{0j} < \cdots < \tilde{C}_{k_j} = +\infty\]
for \( j = n-k+1, \ldots, n \), for which the random vector \((N_1, \ldots, N_k)\);
whose \(i^{th}\) component is given by,

\[
N_i = \sum_{i=n-k+1}^{n} I[C_{i-1,j} < X_j \leq C_{i,j}], \quad i = 1, \ldots, k;
\]

has a multinomial distribution with probabilities \(p_1, \ldots, p_k\);

\[
\sum_{i=1}^{k} N_i = k.
\]

Proof

The proof is an immediate application of Theorem 4.3.4, selecting the transformations in reverse order.

For each \( j = n-k+1, \ldots, n \), let \( Y_j = F_n(X_j | X_1, \ldots, X_{j-1}) \) and define \( C_{i,j} \) by:

\[
F_n(C_{i,j} | X_1, \ldots, X_{j-1}) = \sum_{j=1}^{i} p_j, \quad i = 1, \ldots, k-1.
\]

Note that,

\[
Y_j \in (\sum_{j=1}^{i-1} p_j, \sum_{j=1}^{i} p_j)
\]

iff

\[
C_{i-1,j} < X_j \leq C_{i,j};
\]

thus the result follows from the fact that \(Y_{n-k+1}, \ldots, Y_n\) are distributed independently and uniformly on \((0, 1)\).

Corollary 1

If in the above theorem \(p_j = \frac{1}{k}\), \(j=1, \ldots, k\), then the \(X^2\) statistic,

\[
\sum_{i=1}^{k} \frac{N_i^2}{\bar{N}} - \bar{N} \text{ is distributed as } \chi^2_{\ell,k} \text{ of Section 4.3.4.}
\]
**Corollary 2**

If the minimal sufficient statistic in the above theorem is doubly sequential, the asymptotic distribution of $X^2$ defined in Corollary 1, is that of $X^2_{(k-1)}$.

Theorem 5.2, provides a distribution-free test, which is class-free with respect to the collection $\mathbb{M}_n$ of classes $\mathcal{P}$ whose absolute continuity rank is $\ell$. If the class-freeness notion is relaxed to allow the distribution of the test statistic to vary with $\ell$ within some known family of distributions, then, in this sense Corollary 1 to Theorem 5.2, provides a class-free statistic with respect to the collection $\mathbb{M}$ of classes $\mathcal{P}$ whose absolute continuity rank is positive.

Watson [22] and [23] has proposed the use of randomly selected cells in constructing a chi-square statistic for the composite case. Instead of using MVUE for the distribution, he estimates the parameter $\theta$ (m-dimensional) which identifies the distribution by $\hat{\theta}$, the MLE. The statistic proposed is

$$X^2 = \sum_{i=1}^{k} \frac{n}{F(\hat{C}_i; \hat{\theta}) - F(\hat{C}_{i-1}; \hat{\theta})} \left[ F_n(\hat{C}_i; \hat{\theta}) - F_n(\hat{C}_{i-1}; \hat{\theta}) - F(\hat{C}_i; \hat{\theta}) + F(\hat{C}_{i-1}; \hat{\theta}) \right]^2$$

where $\hat{C}_i$, $i=1, \ldots, k$, is defined by

$$F(\hat{C}_i; \hat{\theta}) = \sum_{j=1}^{i} p_j ,$$

$p_1, \ldots, p_k$ being positive prescribed probabilities.

With the above statistic, it is shown that the asymptotic distribution of $X^2$ is that of,
\[ \chi^2_{(k-m-1)} + \sum_{j=1}^{m} \lambda_j Z_j^2, \]

where \( Z_1, \ldots, Z_m \) are NID (0, 1) and independent of \( \chi^2_{(k-m-1)} \). The numbers \( \lambda_1, \ldots, \lambda_m \) are some eigen values in (0, 1) which are identified. In general, \( \lambda_1, \ldots, \lambda_m \) might depend upon \( \theta \). In the case of \( \theta \) corresponding to location and scale parameters it is shown that \( \lambda_1, \ldots, \lambda_m \) do not depend upon \( \theta \), thus yielding an asymptotically distribution-free statistic.

Moore [15] has extended Watson's results to the multivariate case.
6. APPLICATIONS

In this section the chi-square test proposed in Chapter 5 or, more generally, the random transformations proposed in Chapter 4 are illustrated.

As pointed out in Chapter 4, the random transformations can be obtained also for multivariate problems. The generalized Kolmogorov-Smirnov statistic, however, cannot be generalized to multivariate problems; see Simpson [20].

The examples discussed are straightforward application of the results of Chapter 5 and the MVUE estimators given by Sathe and Varde [18] for the univariate distributions, and by Ghurye and Olkin [7] for the multivariate normal cases.

In Section 6.4, a test for the goodness-of-fit of a univariate multiple regression model is given. In this example, a slightly different setting occurs, in that, the available independent sample is not formed by identically distributed r.v.'s.

6.1 Test for Normality, $\mu$, $\sigma^2$ Unknown

Let

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2.$$ 

For the normal class

$$T_n = (\bar{X}_n, S_n^2)$$

is the minimal sufficient statistic; it is also a doubly sequential statistic.
The conditional distribution based on \( r \) observations, \( \tilde{F}_r(z) \) is given by:

\[
\tilde{F}_r(z) = \begin{cases} 
0, & \text{if } z - \bar{X}_r < -(r-1) \frac{1}{2} s_r, \\
1, & \text{if } (r-1) \frac{1}{2} s_r < z - \bar{X}_r, \text{ and} \\
G_{r-2} \left[ \frac{(r-2) \frac{1}{2} (z-\bar{X}_r)}{[(r-1) s_r^2 - (z - \bar{X}_r)^2]^{\frac{1}{2}}} \right], & \text{elsewhere}
\end{cases}
\]

where \( G_{r-2} \) is the student's t-distribution with \( r-2 \) d.f.

Since \( \tilde{F}_r(z) \) is absolutely continuous for \( r \geq 3 \), then for \( j = n - \ell + 1, \ldots, n \), using the doubly sequentiality property, the transformation for \( x_j \) is given by \( \tilde{F}_j(x_j) \). Thus, \( \ell = n - 2 \).

Let \( \frac{t_{i/k,j-2}}{k} \) be the \( (i/k)^{th} \) percentile from a student's t-distribution with \( j-2 \) d.f.; for \( j = 3, \ldots, n \). Obviously, this selection is for equiprobable cells.

The estimated quantiles

\[
\tilde{c}_{ij}, \quad i = 1, \ldots, k-1,
\]

for the \( j^{th} \) variable are given by the solutions to the equations:

\[
(j-2)^{\frac{1}{2}} \left[ \tilde{c}_{ij} - \bar{X}_j \right] / \left[ (j-1) s_j^2 - (\tilde{c}_{ij} - \bar{X}_j)^2 \right]^{\frac{1}{2}} = \frac{t_{i/k,j-2}}{k},
\]

which gives:

\[
\tilde{c}_{ij}^2 = \frac{((j-1)/(j-2))^{\frac{1}{2}} \cdot \frac{t_{i/k,j-2}}{k} \cdot s_j}{(1 + \frac{t_{i/k,j-2}^2}{k \cdot (j-2)})^{\frac{1}{2}}} + \bar{X}_j.
\]
Note the sequentiaility of the computational procedure, in that \(\bar{X}_j, S_j^2\) is computed first, then \(\bar{X}_h, S_h^2\), etc., and also, if the cells are computed for a sample size \(n_o\), and extra observations are then obtained, the formulas for the first cells are unaltered.

The case of testing normality, when \(\sigma^2\) is known (a case which may be of little practical importance) can be obtained in straightforward manner. The expression for \(\tilde{F}_r(z)\) in this case is that of a \(N(\bar{X}_r, ((r-1)/r)\sigma^2)\).

6.2 Test for Exponentiality

It is desired to test the composite null hypothesis that the parent has density \(\lambda e^{-\lambda t}, \lambda > 0\). The sample mean \(\bar{X}_n\) is the minimal sufficient statistic and is clearly doubly sequential. The Rao-Blackwell estimating distribution function for

\[
\int_0^x \lambda e^{-\lambda t} dt,
\]

is

\[
\tilde{F}_r(z) = 1 - (1 - z/r\cdot\bar{X}_r)^{r-1}
\]

which is defined for \(r \geq 2\).

The absolute continuity rank is thus \(n-1\) and in this case no tables are needed to select the estimated quantiles,

\[
\tilde{C}_{ij}, i = 1, \ldots, k-1, j = 2, \ldots, n.
\]

The \(\tilde{C}_{ij}\) are obtained by solving:

\[
1 - (1 - \tilde{C}_{ij}/j\cdot\bar{X}_j)^{j-1} = \frac{i}{k}
\]
which yields:
\[
\tilde{C}_{1j} = [1 - (1 - \frac{1}{k})^{j-1}] \tilde{X}_j.
\]

Following the same steps as in the above examples, the random selection of cells can be easily obtained for the following cases given by Sathe and Varde, [18]:

Incomplete Gamma; \( \frac{\theta^{-p}}{\Gamma(p)} e^{-\frac{x}{\theta}} x^{p-1} \) for \( p \) known.

Weibull; \( (p/\theta)x^{p-1} e^{-\frac{x^p}{\theta}} \), \( p \) known.

In both cases, the absolute continuity rank is \( n-1 \), and in both cases, also, the corresponding minimal sufficient statistic is doubly sequential.

6.3 Multivariate Normal Test

Consider \( \bar{X}_1, \ldots, \bar{X}_n \) to be a sample of \( k \)-variate independent random vectors from a \( N(\mu, \Sigma) \), where \( \Sigma \) is known.

The minimal sufficient statistic for the above family, is the vector of sample means, \( \bar{X}_n \), which is clearly doubly sequential.

Ghurye and Olkin, give the distribution of any one term in the sample given \( \bar{X}_n \), which is a normal distribution with mean vector \( \bar{X}_n \) and covariance matrix, \( (1 - 1/n)\Sigma \).

Because of the double sequentiality, it is clear that the absolute continuity rank for the family is \( n-1 \).

Let \( j = 2, \ldots, n \) be arbitrary, and consider the sample mean vector \( \bar{X}_j \) computed from the first \( j \) observations, and let
\[ \mathbf{x}_j' = (x_{1j}', x_{2j}', \ldots, x_{kj}') \]

be the \( j \)th observed vector.

From the conditional distribution \( \mathcal{F}_j(x_j') \), given by the

\[ \mathcal{N}(\mathbf{x}_j, (1 - 1/j)V) \]

evaluated at \( \mathbf{x}_j' \), the \( k \) conditional distributions

\[ \mathcal{F}_j(x_{1j}), \mathcal{F}_j(x_{2j}|x_{1j}), \ldots, \mathcal{F}_j(x_{kj}|x_{1j}, \ldots, x_{k-1j}) \]

are obtained as follows:

For any \( s = 2, \ldots, k \), let \( \sigma_s^2 \) be the \( s \)th diagonal element of \( V \), \( V_{s-1} \) be the sub-matrix of \( V \) obtained by considering the first \((s-1)\) rows and columns, and let \( v_{s-1} \) be the vector for which the following relation holds:

\[ V_s = \begin{bmatrix} V_{s-1} & v_{s-1} \\ v_{s-1}' & \sigma_s^2 \end{bmatrix} \]

Let

\[ \mathbf{x}_j' = (\mathbf{x}_{1j}', \ldots, \mathbf{x}_{kj}'); \]

then \( \mathcal{F}_j(x_{1j}) \) is given by a

\[ \mathcal{N}(\mathbf{x}_{1j}', (1 - 1/j)\sigma_1^2) \]

evaluated at \( x_{1j} \); and for \( s = 2, \ldots, k \),

\[ \mathcal{F}_j(x_{sj}|x_{1j}, \ldots, x_{s-1j}) \]
is given by a

\[ N(\tilde{\mu}_{sj}', (1 - 1/j)\sigma^2_s |_1, \ldots, s-l) \]

where

\[ \tilde{\mu}_{sj} = \bar{x}_{sj} + \frac{v'}{s-l} V^{-1}_{s-l}((x_{1j}, \ldots, x_{s-lj})' - (\bar{x}_{1j}, \ldots, \bar{x}_{s-lj})') \]

and,

\[ \sigma^2_s |_1, \ldots, s-l = \sigma^2_s - \frac{v'}{s-l} V^{-1}_{s-l} V_{s-l} \]

These obtained as the usual conditional mean and variance.

If sufficiently accurate tables are available, the transformations can be carried out, obtaining \((n-l)\cdot k\), independent and uniformly distributed random variables in \((0, 1)\). In the case of selecting a chi-square statistic for the transformed observations, the use of only some prescribed percentiles of the normal tables will be needed. In this latter case, suppose \(M\) cells are to be selected with equal probability. The expression for each of the \((n-l)\cdot k\cdot (M-l)\) percentile estimators is given by:

\[ \tilde{C}_{ijs} = \{(1 - 1/j)\sigma^2_s |_1, \ldots, s-l\}^{1/2} z_i + \tilde{\mu}_{sj} \]

where \(z_i\) is the \(\left(\frac{i}{M}\right)\)th percentile of a \(N(0, 1)\), \(i = 1, \ldots, M-1; j = 2, \ldots, n\) and \(s = 1, \ldots, k\).

6.4 Testing the Fit of a Regression Model

Let \(y_1, \ldots, y_n\) be a sample of r.v.'s and let \(y_n' = (y_1, \ldots, y_n)\).

Consider the hypothesis:
(6.4.1) \[ \frac{Y}{n} \sim N(X_n \beta, \sigma^2 I), \]

where \( X_n \) is some fixed \( n \times p \) matrix \((n > p)\), of full rank, and \((\beta, \sigma^2)\) are \( p + 1 \) unknown parameters.

(6.4.1) is the well known univariate multiple regression model.

In this section, a test is given for testing the hypothesis (6.4.1).

For the family \( N(X_n \beta, \sigma^2 I) \), clearly \((y'_{n-n}, X'_{n-n})\) is the minimal sufficient statistic, or equivalently, the statistic,

\[(X'_{n-n}, Y_n') (I - X_n (X'X_n)^{-1}X'_n) Y_n'.\]

Since the family is defined with the knowledge of \( X_n \), then for \( x_i' \), the \( i \)th row of \( X_n \), the statistic:

\[(y'_{s-s}, x'_s y_s, y_{s+1})\]

is known iff

\[(y'_{s+1} y_{s+1}, x'_s y_{s+1}, y_{s+1})\]

is known, which shows that the minimal sufficient statistic is doubly sequential.

Ghurye and Olkin [7], p. 1268, 4.2 gives the MVUE of a related density which in the particular case mentioned above, can be written as follows:

The conditional density of \( y_n \) given \( t_n = X'_n y_n \) and

\[ S_n = y'_n (I - X_n (X'X_n)^{-1}X'_n) Y_n, \]
exists if \( X_{n-1} \) is of full rank, and in that case, it is given by the expression:

\[
\frac{S_n^{n-p-2}}{\beta(n, (n-p-1)/2)} (1 - x_n'(x_n'x_n)^{-1}x_n) \frac{\psi}{\Phi^2} \left[ S_n \left( \frac{(y_n - x_n'(x_n'x_n)^{-1}t_n)^2}{(1 - x_n'(x_n'x_n)^{-1}x_n)} \right) \right]
\]

where,

\[
\psi(z) = \begin{cases} 
  z, & \text{if } z > 0, \\
  0, & \text{if } z \leq 0.
\end{cases}
\]

Therefore, for \( j = p + 2, \ldots, n \), the conditional density of \( y_j \) given \( t_j = x_j'y_j \) and,

\[
S_j = y_j'(I - x_j'(x_j'x_j)^{-1}x_j)y_j
\]

exists if the first \( p+1 \) rows of the matrix \( X_n \) are linearly independent. It is assumed that this is the case. This result enables the writing of the conditional distribution of \( y_j \), given \( t_j \) and \( S_j \) as a student's t-distribution function with \( j - p - 1 \) degrees freedom evaluated at:

\[
U_j = \frac{(j - p - 1)\frac{\psi}{\Phi^2}(y_j - x_j'(x_j'x_j)^{-1}t_j)}{(1 - x_j'(x_j'x_j)^{-1}x_j)S_j - (y_j - x_j'(x_j'x_j)^{-1}t_j)^2)^{\frac{1}{2}}}
\]

Under the assumption that the ordering of the observations is given as above, then the conditional distribution of the \( j^{th} \) observation \( (j = p + 2, \ldots, n) \) given \( S_j \) and \( t_j \); is just, \( G_{j-p-1}(U_j) \), for \( U_j \in (-\infty, \infty) \) or, equivalently for

\[
|y_j - x_j'(x_j'x_j)^{-1}t_j| < (1 - x_j'(x_j'x_j)^{-1}x_j)S_j^{\frac{1}{2}}.
\]
As before, if accurate tables are available to achieve the transformation, any desired statistic can be chosen in the transformed space, using the \( n - p - 1 \) transformed \( U(0, 1) \) independent observations. If such tables are not available or because of some other reason, the chi-square statistic is chosen, the corresponding formulas for the selection of the quantiles \( \tilde{C}_{ij} \); \( i = 1, \ldots, k-1 \), (where \( k \)-equiprobable cells are to be selected), and \( j = p + 2, \ldots, n \); are given by:

\[
\tilde{C}_{ij} = \frac{\{S_j(1 - x_j' (X'X_j)^{-1} x_j)/(j-p-1)\}^{\frac{1}{2}} t_{i/k, j-p-1}}{(1 - t_i^2/k, j-p-1/(j-p-1)^{\frac{1}{2}}) + x_j' (X'X_j)^{-1} x_j}
\]

which clearly particularises to the formulas given for the normal example in Section 1 of this chapter for \( p = 1 \).
7. LIST OF REFERENCES


