*This research was partially supported by the Air Force Office of Scientific Research under Grant #AFOSR 72-2386.

†A dissertation submitted in partial fulfillment for the requirements for the degree of Doctor of Philosophy under the direction of Professor Gordon D. Simons.

STATISTICAL TESTS BASED ON THE LÉVY AND PROHOROV METRICS†

by

Charles H. Alexander*

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 965

November, 1974

CHARLES HAYNES ALEXANDER, JR. Statistical Tests Based on the Lévy and Prohorov Metrics. (Under the direction of GORDON D. SIMONS.)

The author develops non-parametric one-sample and two-sample goodness-of-fit tests using as test statistics the Lévy and Prohorov distance between empirical distribution functions. Computational procedures are described for computing the test statistics. Recurrence equations are described for computing the distribution of the two-sample test statistics, using results about the maximal matchings in certain graphs. The asymptotic distribution of the one-sample test statistic is expressed in terms of the distribution of fluctuations in the sample path of the Brownian Bridge stochastic process. Tables of these distributions are given in the appendix. The power of the tests against certain alternatives is discussed, and the results of simulations comparing the power with that of the Kolmogorov-Smirnov test.

ACKNOWLEDGEMENTS

I wish to thank my advisor, Professor G.D. Simons. His suggestions and encouragement combined with a remarkable aversion to mathematical nonsense have assisted the author immensely throughout. More than an advisor, he has proved to be an educator in the broadest sense of the word and it has been a rewarding experience to work with him.

My thanks also go to the members of the faculty of the Department of Statistics who have contributed so much to my graduate education, particularly the members of my examination committee, Professors Hoeffding, Cambanis, Chakravarti, and Wegman.

I am grateful for the financial support provided by the Department of Statistics and also for the assistance of the Department of Mathematics, especially Professors Cannon and Kelly who have been members of my committee.

Let me here add to the praises of June Maxwell who has ably, speedily, and most of cheerfully typed this manuscript under combat conditions.

Finally, I would like to thank my fellow students for their contribution to my education, particularly N.I. Fisher, whose help in the preparation of this manuscript has been most appreciated.

TABLE OF CONTENTS

Chapter I. Introduction

1.1 Outline and Summary-- 1

1.2 Weak Convergence of Measures------------------------------- 4

1.3 The metrics d_L and d_P --------------------------------- 5

1.4 Asymptotic Distributions Based on the Wiener Process--------12

1.5 Terminology from the Theory of Graphs-----------------------14

Chapter 2. Computational Procedures

2.1 Computation of $d_P(P,Q)$ for General S when P and Q have Finite Support--------------------------------------16

2.2 Computation of $d_P(P,Q)$ when P has Finite Support---------20

2.3 Computation of d_L --24

Chapter 3. Null Distribution for the Two-Sample Test Based on d_P

3.1 Introduction--27

3.2 Matchings in the Graphs (X,Y,Γ_z)----------------------------29

3.3 Definition of a Mapping $a: X \to X$ --------------------------41

3.4 Classification Scheme for Elements of X--------------------43

3.5 Recurrence Equations for $K(u)$, $u \in C$ -----------------------44

3.6 Reducing the Recurrence Equations---------------------------53

3.7 Computating of $\#D_i(l,s)$-----------------------------------55

Chapter 4. Asymptotic Distribution of the Test Statistics

4.1 Asymptotic Distribution for d_P ----------------------------62

4.2 Asymptotic Distribution for d_L ----------------------------74

4.3 Asymptotic Continuity of $f(W^\circ)$----------------------------75

Chapter 5. Statistical Properties

5.1 Related Test Statistics-------------------------------------86

5.2 Power of the Test Based on d_P-------------------------------88

Appendices

A.1 Tables of the One Sample Distribution Function--------------97

A.2 Tables of the Two Sample Distribution Function-------------101

A.3 Bounds on $Pr\{f(W^{\circ}) \leq c\}$ ---------------------------------110

Bibliography---113

CHAPTER 1.

Introduction

1.1 Outline and Summary

Let (S,ρ) be a metric space. Let $\mathcal{S}$ be the class of Borel subsets of S and let $\mathcal{P}(S)$ be the space of all probability measures on $(S,\mathcal{S})$. A number of ways of defining a metric on the space $\mathcal{P}(S)$ have been suggested and it seems natural to use such a metric to construct tests of statistical hypotheses. For a one-sample goodness-of-fit test, the metric may be used to measure the distance between the empirical measure of a random sample and the probability measure from which the sample is hypothesized to have been drawn. (By "empirical measure" is meant the measure which puts equal mass at all the sample points and zero mass elsewhere.) The corresponding two-sample test for identical distribution would use as a test statistic the distance between the empirical measures of the two samples. The standard example is the Kolmogorov-Smirnov test, where S is the real line and the uniform metric on $\mathcal{P}(S)$ is used, i.e., the metric $d_{\mathcal{K}}(P,Q) = \sup_{-\infty<t<\infty} |P(-\infty,t] - Q(-\infty,t]|$. The Cramér-von Mises test is based on a directed distance between the empirical distribution function F_n and the null distribution function F, $d_{\mathcal{C}}(F_n,F) = n\int_{-\infty}^{\infty} (F_n(x)-F(x))^2\Psi(F(x))dF(x)$, where $\Psi(\cdot)$ is a weight function. ($d_{\mathcal{C}}$ is not metric since it is not symmetric in its arguments.) Other tests based on metrics on $\mathcal{P}(S)$ have been suggested in [1] and [26].

The so-called "Levy" metric, d_L , which is defined for $S = E^k$ (k-dimensional Euclidean space) and has as its metric topology the topology of weak convergence of probability measures, and the "Prohorov" metric, d_P , which generates the same topology on $P(S)$ for a more general class of metric spaces S, have been given little discussion in the context of hypothesis testing. Dudley [12] remarks that there exists no practical means for computing the value of the test statistic for the test based on d_P. Zolotarev [32] suggests that a barrier to the use of tests based on d_L is the lack of knowledge about its characteristic function, which presumably prevents the development of the asymptotic distribution theory along the lines of that for the Kolmogorov-Smirnov test.

An additional difficulty is that while the statistical tests based on d_K are distribution-free for continuous null distributions, the analogous tests based on d_P and d_L are not. In the case $S = E^1$, we can construct a distribution-free one-sample test by using d_P or d_L to measure the distance between the uniform measure on [0,1] and the empirical measure of $\{F(X_i)\}$, where F is the null distribution function and $\{X_i\}$ denotes the sample. The corresponding two-sample test would use d_P or d_L to measure the distance between the empirical measures of $\{s_i/N\}$ and $\{r_i/N\}$, where $r_1,\ldots,r_m$ and $s_1,\ldots,s_n$ are the respective ranks of the two samples among the pooled observations, and $N = m+n$. Alternatively, the two-sample test statistic could be the distance between the empirical measure of $\{s_i/N\}$ and the empirical measure of $\{1/N,\ldots,N/N\}$.

In what follows, we will discuss the computation of these test statistics based on d_P and d_L, their distributions under the null hypothesis, and some of their statistical properties. For d_P, exact distributions of the two-sample test statistic are computed for the case of equal sample sizes, corresponding to the results of Massey [21] for the Kolmogorov-Smirnov test. We are unable to give similar results for the one-sample test, as are found by Birnbaum and Tingey [6] for the Kolmogorov-Smirnov test. The asymptotic distributions of the test statistics are related to the distribution of particular fluctuations in the sample path of the "Brownian Bridge" process. As with d_K, there is a single asymptotic distribution for the one-sample and two-sample test statistics. We have not succeeded in finding a closed-form expression for this distribution, although some crude upper and lower bounds have been computed. The asymptotic distribution of the test statistics based on d_L is closely related to that of the statistics based on d_K, and can be given explicitly.

In Chapter 1, we give definitions and a brief description of weak convergence and of the metrics d_L and d_P. Results and relationships are collected here which will be of use in later chapters. Chapter 2 contains computational procedures for d_P and d_L. One algorithm computes $d_P(P,Q)$ for measures P and Q which each have finite support. For $S = E^1$, a more efficient algorithm computes $d_P(P,Q)$ where P has finite support and Q is arbitrary. In Chapter 3, recurrence equations are derived which are used to compute the exact null distribution of the two-sample test statistic for equal sample sizes. Chapter 4 contains the results on the asymptotic distributions of the test statistics.

Chapter 5 consists of a brief discussion of the statistical properties of the tests, including the results of simulations comparing the power of these tests with the power of the Kolmogorov-Smirnov test against normal location and scale alternatives, and against "Lehmann" nonparametric alternatives. An appendix contains tables of the distribution of the two-sample test statistic obtained using the methods of Chapter 3, and the results of Monte Carlo estimation of the exact distribution of the one-sample test statistic for sample sizes 5, 10, 20, 40, 60, and 80, as well as a description of the computer programming methods used.

1.2 Weak convergence of measures.

As before, let (S,ρ) be a metric space and let $\mathcal{S}$ be the σ-field of all Borel subsets of S. Let $\mathcal{P}(S)$ be the space of all probability measures on $(S,\mathcal{S})$. Let $\{P_n\}$ and P be elements of $\mathcal{P}(S)$.

Definition 1.2.1: (a) $\{P_n\}$ is said to *converge weakly* to P (written $P_n \Rightarrow P$) if and only if $\int_S f\, dP_n \to \int_S f\, dP$ for every bounded continuous real-valued function f defined on S.

(b) If $S = E^k$ and $F(x) = P\{y: y \le x\}$ and $F_n(x) = P_n\{y: y \le x\}$, $n=1,2,\ldots,$ then we say that $F_n \Rightarrow F$ if and only if $P_n \Rightarrow P$. (Note: if $x = (x_1,\ldots,x_k)$ and $y = (y_1,\ldots,y_k)$, we write $y \le x$ if and only if $y_i \le x_i$, $i=1,\ldots,k$.)

(c) Let T be a metric space and $\mathcal{T}$ the Borel subsets of T. If $\{X_n\}$ and X are random elements (i.e., measurable mappings) from a probability space $(\Omega,\mathcal{A},P)$ to $(T,\mathcal{T})$, we say that $X_n \Rightarrow X$ if and only if $PX_n^{-1} \Rightarrow PX^{-1}$ as probability measures in $\mathcal{P}(T)$. In this case we say that $\{X_n\}$ converges in distribution to X. $\square$

Definition 1.2.2: For each $A \subseteq S$, let δA denote the boundary of A. Let $P \in \mathcal{P}(S)$. $A \in \mathcal{S}$ is called a P-*continuity set* if and only if $P(\delta A) = 0$. (Note that $\delta A \in \mathcal{S}$, since δA is closed.) □

We will usually use the following equivalent characterization of weak convergence instead of the original definition. The next two Propositions are found in [5], where additional equivalent conditions for weak convergence are given.

Proposition 1.2.3: $P_n \Rightarrow P$ if and only if $P_n(A) \to P(A)$ for all P-continuity sets A. If $S = E^k$ and $\{F_n\}$ and F are defined as before, then $F_n \Rightarrow F$ if and only if $F_n(x) \to F(x)$ for every point $x \in E^k$ such that F is continuous at x. □

Proposition 1.2.4: Let T be a separable metric space and $\mathcal{T}$ the Borel σ-field of T. Let $\{h_n\}$ and h be measurable functions from $(S,\mathcal{S})$ to $(T,\mathcal{T})$. Let E be the set of all $x \in S$ for which there exists some sequence $\{x_n\}$ of points of S such that $x_n \to x$ but $h_n(x_n) \not\to h(x)$. (Since T is separable, $E \in \mathcal{S}$.) If $P_n \Rightarrow P$ and $P(E) = 0$, then $P_n h_n^{-1} \Rightarrow Ph^{-1}$. □

Proposition 1.2.4 frequently appears through its corollary that if h is a continuous function from $(S,\mathcal{S})$ to $(T,\mathcal{T})$, then $P_n \Rightarrow P$ implies $P_n h^{-1} \Rightarrow Ph^{-1}$.

1.3 The metrics d_L and d_P.

Definition 1.3.1: (a) Let $S = E^k$. Define the metric d_K on $\mathcal{P}(S)$ by $d_K(P,Q) = \inf\{\varepsilon>0: P(-\infty,x] - \varepsilon \le Q(-\infty,x] \le P(-\infty,x] + \varepsilon$;

for all $x \in E^k\}$.

(b) Let $S = E^k$. Define the metric d_L on $\mathcal{P}(S)$ by

$d_L(P,Q) = \inf\{\varepsilon>0\colon P(-\infty,x-\varepsilon] - \varepsilon \le Q(-\infty,x] \le P(-\infty,x+\varepsilon] + \varepsilon$; for all $x \in E^k\}$.

(c) Let (S,ρ) be an arbitrary metric space. Define the metric $d_{\mathcal{P}}$ on $\mathcal{P}(S)$ by $d_{\mathcal{P}}(P,Q) = \inf\{\varepsilon>0\colon Q(F) \le P(F^\varepsilon) + \varepsilon$; for all closed sets $F \subseteq S\}$, where $F^\varepsilon = \{y \in S\colon$ there exists $x \in F$ such that $\rho(x,y) < \varepsilon\}$.

□

(Note: the definition of $d_{\mathcal{P}}$ is not apparently symmetric in P and Q. However, suppose that $Q(F) \le P(F^\varepsilon) + \varepsilon$ for all closed sets $F \subseteq S$. Let $G \subseteq S$ be closed, then G^ε is open. Since $G \subseteq S - (S-G^\varepsilon)^\varepsilon \subseteq G^\varepsilon$ and $Q(S-G^\varepsilon) \le P((S-G^\varepsilon)^\varepsilon) + \varepsilon$, we have the inequalities $P(G) \le P(S-(S-G^\varepsilon)^\varepsilon) \le Q(G^\varepsilon) + \varepsilon$, hence $d_{\mathcal{P}}(P,Q) = d_{\mathcal{P}}(Q,P)$. This proof appears in [30]. Previously the definition of the Prohorov distance between P and Q was usually given as what, in our notation would be, $\min(d_{\mathcal{P}}(P,Q),\ d_{\mathcal{P}}(Q,P))$. This latter definition is more general since when dealing with a larger space of measures so that $P(S) \ne Q(S)$, $d_{\mathcal{P}}(P,Q)$ may not equal $d_{\mathcal{P}}(Q,P)$.)

<u>Proposition 1.3.2</u>: (a) Let S be separable and complete. $P_n \Rightarrow P$ if and only if $d_{\mathcal{P}}(P_n,P) \to 0$.

(b) If $S = E^k$, $P_n \Rightarrow P$ if and only if $d_L(P_n,P) \to 0$.

The proof of (a) first appears in [24]. A proof of (b) for k = 1 is given in [14]. Bhattacharya in [3] observes that essentially the same proof works for $k > 1$. Prohorov [24] shows that the metric space $(\mathcal{P}(S),d_{\mathcal{P}})$ is separable and complete when S is separable and complete.

These assumptions about S can be weakened (see [5]). □

1.3.2 shows that d_P and d_L are equivalent metrics, i.e., they generate the same topology on $P(S)$, when $S = E^k$. However, they do not generate the same uniformities, as shown by the following example from [11]. Define P_n and Q_n by $P_n(2j) = Q_n(2j+1) = 1/n$, $j = 1,\dots,n$. Then $d_L(P_n,Q_n) = 1/n$, while $d_P(P_n,Q_n) = 1/2$. The topologies generated by these metrics are relevant to their statistical properties. Two examples to this effect are suggested by Hampel's remark in [15] that d_P gives a "literal quantitative description" of "rounding of the observations" and "the occurrence of gross errors."

Example 1.3.3: Let $x_1,\dots,x_n$ be real numbers and define $y_1 = x_1 + \varepsilon$, and $y_i = x_i$, for $i = 2,3,\dots,n$. Let P_n and Q_n be the empirical measures of $\{x_i\}$ and $\{y_i\}$ respectively. Assume that $0 < \varepsilon < \min_{i \neq j}|x_i - x_j|$. Then $d_P(P_n,Q_n) = d_L(P_n,Q_n) = \min(\varepsilon,1/n)$, while $d_K(P_n,Q_n) = 1/n$. In this case, when $\varepsilon < 1/n$, d_L and d_P give a better idea of the size of the "error" in the first observation. □

Example 1.3.4: Now define $y_i = x_i + \varepsilon$, $i = 1,\dots,k$, $y_i = x_i$ for $i = k+1,\dots,n$. Let $\varepsilon > k/n$. Then $d_P(P_n,Q_n) = k/n$, while $d_L(P_n,Q_n) = 1/n$. □

Neither d_L nor d_P generates a distribution-free test. In particular, neither metric is invariant under transformations corresponding to changes of scale in the data. This lack of invariance must hold for any metric which generates the topology of weak convergence on $P(E^k)$.

The following Glivenko-Cantelli-type result is due to Varadarajan [31].

Proposition 1.3.5: Let $P \in \mathcal{P}(S)$, where S is separable and complete, and let P_n be the empirical measure of a sample of size n from P. Then $P_n \Rightarrow P$ a.s. Hence $d_{\mathcal{P}}(P_n,P) \to 0$ a.s., and for $S = E^k$, $d_{\mathcal{L}}(P_n,P) \to 0$ a.s. □

The definitions which we have given for $d_{\mathcal{P}}$ and $d_{\mathcal{L}}$ are those most often found in the literature. They are actually not the most convenient for our purposes and the following results show that certain modifications can be made. Let $F^{\varepsilon]} = \{y \in S: x \in F$ such that $\rho(x,y) \le \varepsilon\}$.

Proposition 1.3.6: (a) Let $P, Q \in \mathcal{P}(S)$. Then $d_{\mathcal{P}}(P,Q) = \inf\{\varepsilon > 0: Q(F) \le P(F^{\varepsilon]}) + \varepsilon;$ for all closed $F \subseteq S\}$.
(b) Let $P, Q \in \mathcal{P}(E^k)$. Then $d_{\mathcal{L}}(P,Q) = \inf\{\varepsilon > 0: P(-\infty,x-\varepsilon) - \varepsilon \le Q(-\infty,x] \le P(-\infty,x+\varepsilon] + \varepsilon;$ for all $x \in E^k\}$.

Proof: (a) For each closed $F \subseteq S$,

$P(F^{\varepsilon}) + \varepsilon \le P(F^{\varepsilon]}) + \varepsilon \le P(F^{\varepsilon'}) + \varepsilon'$, for each $\varepsilon > 0$ and $\varepsilon' > \varepsilon$.
(b) For each $x \in E^k$,
$P(-\infty,x-\varepsilon'] - \varepsilon' \le P(-\infty,x-\varepsilon) - \varepsilon \le P(-\infty,x-\varepsilon] - \varepsilon$, for each $\varepsilon > 0$ and $\varepsilon' > \varepsilon$. □

Definition 1.3.7: For given $P, Q \in \mathcal{P}(S)$, define the function $V_{\mathcal{P}}(P,Q;\varepsilon)$ of ε by $V_{\mathcal{P}}(P,Q;\varepsilon) = \sup(P(F) - Q(F^{\varepsilon]}))$, the supremum being taken over all closed $F \subseteq S$. Define $V_{\mathcal{L}}(P,Q;\varepsilon) =$

$$\max\{ \sup_{x\in E^k}(P(-\infty,x-\varepsilon) - Q(-\infty,x]),\ \sup_{x\in E^k}(Q(-\infty,x] - P(-\infty,x+\varepsilon])\}.$$

□

Proposition 1.3.8: (a) For $c \in [0,1]$, $d_P(P,Q) \le c$ if and only if $V_P(P,Q;c) \le c$.

(b) For $c \in [0,1]$, $d_L(P,Q) \le c$ if and only if $V_L(P,Q;c) \le c$.

Proof: (a) By 1.3.6 (a), if $V_P(P,Q;c) \le c$, then $d_P(P,Q) \le c$. Conversely, $d_P(P,Q) \le c$ implies for $c' > c$ that $P(F) \le Q(F^{c']}) + c'$, for all closed $F \subseteq S$. But for each closed set F, $\lim_{c' \downarrow c} Q(F^{c']}) = Q(F^{c]})$, so $P(F) \le Q(F^{c]}) + c$ for all closed $F \subseteq S$, i.e., $V_P(P,Q;c) \le c$.

(b) Similarly, $\lim_{c' \downarrow c} P(-\infty, x-c') = P(-\infty, x-c)$ and $\lim_{c' \downarrow c} P(-\infty, x+c'] = P(-\infty, x+c]$. □

The next two results show that when $S = E^1$ or $S = [0,1]$ under the Euclidean metric, then in defining d_P it is not necessary to consider all closed subsets of S, but that a smaller class of sets will suffice.

Definition 1.3.9: Let $\mathcal{F}$ be the class of all closed subsets of $[0,1]$ which can be expressed as the union of finitely many disjoint closed intervals. Let $\mathcal{F}^{(2\delta)}$ be the class of all members of $\mathcal{F}$ such that if we write F as the union of the disjoint intervals $I_j = [b_j, c_j]$, $j = 1,\ldots,m$, where $0 \le b_1 \le c_1 < b_2 \le c_2 < \ldots < b_m \le c_m \le 1$, then $b_{i+1} - c_i > 2\delta$ for $i = 1,\ldots,m-1$; and if $b_1 \ne 0$, then $b_1 > \delta$; and if $c_m \ne 1$, then $c_m < 1 - \delta$. □

Proposition 1.3.10: Let $P, Q \in \mathcal{P}([0,1])$. Then $d_P(P,Q) \le c$ if and only if $\sup_{G \in \mathcal{F}^{(2c)}} (P(G) - Q(G^{c]})) \le c$.

Proof: Let $\varepsilon > 0$ be arbitrary. It suffices to show that $V_P(P,Q,c) = \sup_{G \in F^{(2c)}} (P(G) - Q(G^{c]}))$. For each closed set $F \subseteq [0,1]$, let the set $G(F)$ be the union of F with every interval $[c_i, b_{i+1}]$ such that $b_{i+1} - c_i \leq 2\delta$, and with $[0,b_1]$ if $b_1 \leq \delta$, and with $[c_m, 1]$ if $c_m \geq 1 - \delta$; where $\{b_i\}$ and $\{c_i\}$ are as in Definition 1.3.9. Then $(G(F))^{c]} = F^{c]}$ and $F \subseteq G(F)$, so $P(F) - Q(F^{c]}) \leq P(G) - Q(G^{c]})$. But $G(F) \in F^{(2c)}$. □

The proof of the same result for $P,Q \in P(E^1)$ is similar except that the sets of F are allowed to consist of the unions of finite or semi-infinite intervals. As immediate corollaries to 1.3.10 it follows that $d_P(P,Q) \leq c$ if and only if $\sup_{F \in F}(P(F) - Q(F^{c]})) \leq c$, and that $d_P(P,Q) =$

$$\inf\{\varepsilon > 0: P(F) \leq Q(F^{\varepsilon]}) + \varepsilon \text{ for all } F \in F\} =$$
$$\inf\{\varepsilon > 0: P(F) \leq Q(F^{\varepsilon]}) + \varepsilon \text{ for all } F \in F^{(2\varepsilon)}\} .$$

It follows from Definition 1.3.1 (and the fact that $d_P(P,Q) = d_P(Q,P)$) that for $P,Q \in P(E^k)$, $d_K(P,Q) \leq d_L(P,Q)$ and also that $d_P(P,Q) \leq d_L(P,Q)$, since the inequalities which must, in the definition of d_P, be satisfied for all closed sets F need only be satisfied for sets of the form $(-\infty,x]$ or $[x,\infty)$ in the definition of d_L. No such general inequality exists between d_K and d_P. However, when dealing with measures in $P([0,1])$ and when $P = U$, the uniform measure, the following relationships hold.

Proposition 1.3.11: Let $S = [0,1]$ and $\rho(x,y) = |x-y|$. Let U be Lebesgue measure restricted to S. Let $V \in P(S)$ be arbitrary. Then

$$d_K(U,V)/2 \leq d_P(U,V) \leq d_K(U,V) .$$

Proof: If for some $x \in [0,1]$, $|U[0,x] - V[0,x]| \geq 2\varepsilon > 0$, then either

$$V([0,x]) \geq U([0,x]^{\varepsilon}) + \varepsilon \ , \quad \text{or} \quad V([x,1]) \geq U([x,1]^{\varepsilon}) + \varepsilon,$$

and hence $d_P(U,V) \geq \varepsilon$. Therefore, $d_P(U,V) \geq d_K(U,V)/2$.

To show $d_P(U,V) \leq d_K(U,V)$, suppose $0 < \varepsilon < d_P(U,V)$. Then we may choose a closed set F such that $V(F) > U(F^{\varepsilon}) + \varepsilon$. We may choose F such that $F \in \mathcal{F}^{(2\varepsilon)}$, as in Proposition 1.3.10. Write F as the union of the disjoint intervals $I_j = [b_j,c_j]$, $j=1,\ldots,m$, where $0 \leq b_1 \leq c_1 < b_2 \leq c_2 < \ldots < b_m \leq c_m \leq 1$.

We have one of three cases

(i) $\varepsilon < b_1$ and $c_m < 1 - \varepsilon$;

(ii) $\varepsilon < b_1$ and $c_m = 1$, or $b_1 = 0$ and $c_m < 1 - \varepsilon$

(iii) $0 = b_1$ and $1 = c_m$.

Then $U(F^{\varepsilon}) = U(F) + \lambda\varepsilon$, where $\lambda = 2m$ in case (i); $\lambda = 2m-1$ in case (ii); $\lambda = 2m-2$ in case (ii); and hence

$$\sum_{j=1}^{m} V(I_j) \geq U(F^{\varepsilon}) + \varepsilon = (\lambda+1)\varepsilon + \sum_{j=1}^{m} U(I_j) \ .$$

Suppose case (i) holds. Then for some j, $1 \leq j \leq m$, $V(I_j) > U(I_j) + 2\varepsilon$ and hence $2\varepsilon < (V-U)(I_j) \leq 2d_K(U,V)$.

Suppose case (ii) holds and for definiteness that $c_m = 1$. Then either $V([b_m,1]) \geq U([b_m,1]) + \varepsilon$, so that $d_K(U,V) \geq \varepsilon$, or

$$\sum_{j=1}^{m-1} V(I_j) \geq 2(m-1)\varepsilon + \sum_{j=1}^{m-1} U(I_j) \quad ,$$

and hence for some j, $1 \leq j \leq m-1$, $V(I_j) \geq U(I_j) + 2\varepsilon$, so that $2d_K(U,V) \geq 2\varepsilon$. Suppose case (iii) holds. Then either $V([0,c_1]) \geq U([0,c_1]) + \varepsilon$ or $V([b_m,1]) \geq U([b_m,1]) + \varepsilon$, so that $d_K(U,V) \geq \varepsilon$, or

$$\sum_{j=2}^{m-1} V(I_j) \geq 2(m-2)\varepsilon + \sum_{j=2}^{m-1} U(I_j) \quad ,$$

and hence for some j, $2 \le j \le m-1$, $V(I_j) \ge U(I_j) + 2\varepsilon$ so that $2d_K(U,V) \ge 2\varepsilon$. In all three cases $\varepsilon \le d_K(U,V)$. Letting $\varepsilon \uparrow d_P(U,V)$, the proof is complete. □

This proof is a modification of that of Dudley in [12], where it is shown that $d_K(U,V)/2 \le d_P(U,V) \le 2d_K(U,V)$. The bounds given are at least very nearly the best possible. Indeed, if V is the measure in $P([0,1])$ which places unit mass at 0, then $d_K(U,V) = 1$, while $d_P(U,V) = 1/2$. Let $k \ge 2$. If $V^{(k)}$ places mass $1/2k$ at the points 0 and 1 and mass $1/k$ at $1/k, 2/k, \ldots, (k-1)/k$, then $d_P(V^{(k)},U) = 1/(2k+1)$, while $d_K(V^{(k)},U) = 1/2k$. Hence as $k \to \infty$, $d_P(U,V^{(k)})/d_K(U,V^{(k)}) \to 1$. We have no example in which $d_P(U,V) = d_K(U,V) > 0$.

1.3.11 shows that if U_n is the empirical measure of a sample of size n from U, then $n^{\frac{1}{2}}E(d_P(U_n,U))$ is bounded and bounded away from zero. Dudley [12] also gives asymptotic upper bounds on $E(d_P(V_n,V))$, where $V \in P(S)$, for a more general class of metric spaces S, where V_n is the empirical measure of a sample of size n from V.

1.4 Asymptotic distributions based on the Wiener process.

Definition 1.4.1: (a) The probability measure W defined on $C[0,1]$ such that for $t \in [0,1]$, $W(t)$ is normally distributed with $E(W(t)) = 0$ and $\mathrm{var}(W(t)) = t$, $(W(0) = 0$ a.s.) and such that for $0 \le t_0 \le t_1 \le \ldots \le t_k \le 1$, the random variables $W(t_1) - W(t_0)$, $W(t_2) - W(t_1), \ldots, W(t_k) - W(t_{k-1})$ are independent, will be called *Wiener measure*. It follows that $E(W(t)W(s)) = \min(s,t)$, for $s,t \in [0,1]$.

(b) The measure W° on $C[0,1]$ with $W^\circ(t)$ normal, $E(W^\circ(t)) = 0$, and $E(W^\circ(t)W^\circ(s)) = s(1-t)$ for $s \leq t$ will be called the *Brownian Bridge*. We may define W° by $W^\circ(t) = W(t) - tW(1)$, $0 \leq t \leq 1$. (Alternatively $W(t)$ may be defined for $t \in [0,\infty)$ and then $W^\circ(s)$ is defined as $(1-s)W(s/(1-s))$, $0 \leq s \leq 1$.) The "sample path" $W(t)$ is a continuous function on $[0,1]$, and hence $W^\circ(t)$ is also. $W^\circ(t)$ may also be thought of as being distributed as $W(t)$ conditioned on the event $\{W(1) = 0\}$.

W° has an important role in deriving the asymptotic distribution for a number of test statistics, including the Kolmogorov-Smirnov, making use of Proposition 1.2.4. This was first recognized by Doob [10], and the proof was made rigorous by Donsker [9]. Other examples of the application of this method are [4] and [27]. The following results are found in [5] and [10].

<u>Proposition 1.4.2</u>: (a) Let U be the uniform measure on $[0,1]$ and let U_n be the empirical measure of a random sample of size n from U. Then the process X_n defined by $X_n(t) = n^{\frac{1}{2}}(U_n[0,t] - U[0,t])$, $0 \leq t \leq 1$, converges weakly to W° in the Skorohod topology on $D[0,1]$, the space of all functions on $[0,1]$ which are continuous from the right and have left-hand limits. In particular this means by Proposition 1.2.4 that $n^{\frac{1}{2}}d_K(U_n,U) = \sup_{0\leq t\leq 1} |X_n(t)|$ converges in distribution to $\sup_{0\leq t\leq 1} |W^\circ(t)|$.

(b) Let U_m and V_n be the empirical measures of $\{s_i/N\}$ and $\{r_i/N\}$, where $\{s_i\}$ and $\{r_i\}$ are the ranks among the pooled observations of two random samples from the same continuous distribution and $N = m+n$, m and n being the respective sample sizes. Then as m and n approach infinity, the process $Y_{m,n}$ defined by $Y_{m,n}(t) = (mn/N)^{\frac{1}{2}}(U_m([0,t]) - V_n([0,t]))$ converges weakly to $W°$, provided $0 < a \leq m/n \leq b < \infty$. □

<u>Proposition 1.4.3</u>: Let $M° = \sup_{0\leq t\leq 1} W°(t)$ and $m° = \inf_{0\leq t\leq 1} W°(t)$.

(a) $\Pr\{M° < b\} = 1 - \exp(-2b^2)$, $b > 0$.

(b) $\Pr\{\sup_{0\leq t\leq 1} |W°(t)| \leq b\} = 1 + 2\sum_{k=1}^{\infty} (-1)^k \exp(-2k^2b^2)$, $b > 0$.

(c) $\Pr\{a < m° \leq M° \leq b\} = \sum_{k=-\infty}^{\infty} \exp(-2k^2c^2) - \sum_{k=-\infty}^{\infty} \exp(-2(b+kc)^2)$,

where $c = b-a$, $b > 0$ and $a < 0$. □

These results agree with earlier results proved by Kolmogorov and Smirnov. Billingsley proves (c) by arguments about a sequence of random walks which approximate the process $W°$. Doob uses a more general result about the probabilities of various level crossings by the $W°$ process, using the Inclusion-Exclusion principle to obtain the expression given for $\Pr\{a < m° \leq M° \leq b\}$.

1.5 <u>Terminology from the Theory of Graphs</u>. Let X and Y be finite sets. A *finite simple graph*, Γ, is a collection of *arcs*, or ordered pairs (x,y), with *initial vertex* $x \in X$ and *terminal vertex* $y \in Y$. A *matching* is a subset of Γ in which each element of X (resp. Y) appears at most once as an initial (resp. terminal) vertex. A *maximal matching* is a matching of maximum possible cardinality; there is not in general a unique maximal matching. The number of arcs in a maximal matching in the graph (X,Y,Γ) will be denoted $N(X,Y,\Gamma)$. The *deficiency* of the graph (X,Y,Γ) is defined as $D(X,Y,\Gamma) = \max_{F \subseteq X}(\#(F) - \#(\Gamma(F)))$, where $\Gamma(F) = \{y \in Y$: there exist $x \in F$ such that $(x,y) \in \Gamma\}$, and for any set S, $\#(S)$ denotes the cardinality of S.

Proposition 1.5.1: $N(X,Y,\Gamma) = \#(X) - D(X,Y,\Gamma)$. This result is found in [23]. □

CHAPTER 2

Computational Procedures

2.1 Computation of $d_P(P,Q)$ for general S when P and Q have finite support.

By Proposition 1.3.8(a), we know that $d_P(P,Q) \le c$ if and only if $V_P(P,Q;c) \le c$. For purposes of testing statistical hypotheses, we need only to check whether a single inequality $d_P(P,Q) \le c_\alpha$ holds, where c_α is the critical value of a test of size α. If instead we wish to compute the actual value of $d_P(P,Q)$ to any desired accuracy, we first test whether $d_P(P,Q) \le 1/2$, then, depending on the result, whether $d_P(P,Q) \le 1/4$ or $d_P(P,Q) \le 3/4$, etc., until finally it is determined that $d_P \in (i/2^k, (i+1)/2^k]$ for some value of i between 0 and 2^k-1, and k chosen large enough to provide the desired accuracy. It is therefore our purpose to compute $V_P(P,Q;c)$ for any given value of c.

If P has finite support, i.e., $P(X) = 1$, where $X = \{x_1,\ldots,x_n\}$, then it is clear that $V_P(P,Q;c) = \sup_{F\subseteq X}(P(F) - Q(F^{c]}))$. This supremum is a maximum, so there exists a subset $G \subseteq X$ such that $P(G) - Q(G^{c]}) \ge P(F) - Q(F^{c]})$ for all $F \subseteq S$. When $S = E^1$, we will be able to determine the set G without placing further restrictions on the measure Q. For general S, we are able to solve the problem using Proposition 1.5.1, but only when Q also has finite support.

Let P_n put mass 1/n at each of the points $x_1,\ldots,x_n$ and let Q_n put mass 1/n at each point of $Y = \{y_1,\ldots,y_n\}$. Define the graph (X,Y,Γ_c) by letting the arc $(x_i,y_j) \in \Gamma_c$ if and only if $\rho(x_i,y_j) \le c$, where ρ is the metric on S. Then $V_{\mathcal{P}}(P_n,Q_n;c)$ is just 1/n times the deficiency of (X,Y,Γ_c), and hence by Proposition 1.5.1, $V_{\mathcal{P}}(P_n,Q_n;c) = n^{-1}(n-N(X,Y,\Gamma_c))$. There does not seem to be a method for finding the number of arcs in a maximal matching in a graph except by constructing a maximal matching. There are published algorithms for finding a maximal matching in any graph. The standard method has been the "Hungarian algorithm" (see [2] or [18]). A new and more efficient algorithm has been proposed by Hopcroft and Karp ([16]).

X and Y need not have the same cardinality. Let P_m put mass 1/m at each of the points $x_1,\ldots,x_m$ and let Q_n put mass 1/n at each of the points $y_1,\ldots,y_n$. Then $V_{\mathcal{P}}(P_m,Q_n;c) = \sup_{F\subseteq X}(m^{-1}\#(F) - n^{-1}\#(\Gamma_c(F)))$, where the graph (X,Y,Γ_c) is defined as in the case m = n. Suppose m and n have greatest common factor q, so that $m = p_1q$ and $n = p_2q$, where p_1 and p_2 are relatively prime. Then

$$p_1p_2qV_{\mathcal{P}}(P_m,Q_n;c) = \sup_{F\subseteq X}(p_2\#(F) - p_1\#(\Gamma_c(F))) .$$

Define the graph (X',Y',Γ_c') as follows. X' and Y' each consist of p_1p_2q vertices,

$$X' = \{x_{11},x_{12},\ldots,x_{1p_2},x_{21},x_{22},\ldots,x_{2p_2},\ldots,x_{m1},\ldots,x_{mp_2}\}$$

and

$$Y' = \{y_{11},y_{22},\ldots,y_{1p_1},y_{21},y_{22},\ldots,y_{2p_1},\ldots,y_{n1},\ldots,y_{np_1}\} .$$

Define Γ_c' as follows. For each i and j, $1 \le i \le m$ and $1 \le j \le n$, let $(x_{ir},y_{js}) \in \Gamma'_c$ for all values of r and s, $1 \le r \le p_2$ and $1 \le s \le p_1$, if and only if $(x_i,y_j) \in \Gamma_c$. Intuitively, X' consists of p_2 "copies" of each point of X and Y' consists of p_1 copies of each point of Y, and the "copies" are connected by an arc in Γ'_c, whenever the "originals" are connected by an arc in Γ_c.

<u>Proposition 2.1.1</u>: $\underset{G\subseteq X'}{\mathrm{Max}}(\#(G) - \#(\Gamma'_c(G))) = \underset{F\subseteq X}{\max}(p_2\#(F) - p_1\#(\Gamma_c(F)))$.

<u>Proof</u>: Suppose $G_0 \subseteq X$ and $\#(G_0) - \#(\Gamma'_c(G_0)) = \underset{G\subseteq X'}{\max}(\#(G) - \#(\Gamma'_c(G)))$.

If G_0 contains any "copy" of x_i, then G_0 may be assumed to contain all "copies" of x_i, i.e., if $\{x_{i1},\ldots,x_{ip_2}\} \cap G_0 \neq \emptyset$, then $\{x_{i1},\ldots,x_{ip_2}\} \subseteq G_0$. To see this observe that for each r, $1 \le r \le p_2$, $\Gamma'_c(\{x_{i1},\ldots,x_{ip_2}\}) = \Gamma'_c(\{x_{ir}\})$ and hence, if $x_{ir} \in G_0$,

$$\#(G_0 \cup \{x_{i1},\ldots,x_{ip_2}\}) - \#(\Gamma'_c(G_0 \cup \{x_{i1},\ldots,x_{ip_2}\}))$$
$$\ge \#(G_0) - \#(\Gamma'_c(G_0)) .$$

Let $\mathcal{G}$ be the class of all sets $G \subseteq X'$ such that for each i, $\{x_{i1},\ldots,x_{ip_2}\} \cap G \neq \emptyset$ if and only if $\{x_{i1},\ldots,x_{ip_2}\} \subseteq G$. Then we have just shown that

$$\max_{G\subseteq X'}(\#(G) - \#(\Gamma'_c(G))) = \max_{G\in\mathcal{G}}(\#(G) - \#(\Gamma'_c(G))) .$$

For each set $F^* \subseteq X$, let $G^* = \bigcup_{\{i:x_i\in F^*\}} \{x_{i1},\ldots,x_{ip_2}\}$, in this way defining a 1-1 correspondence between $\mathcal{G}$ and the class of all subsets of X. That $\#(G^*) - \#(\Gamma'_c(G^*)) = p_2\#(F^*) - p_1\#(\Gamma_c(F^*))$ follows from the definition of Γ'. Hence

$$\max_{G\in\mathcal{G}}(\#(G) - \#(\Gamma'_c(G))) = \max_{F\subseteq X}(p_2\#(F) - p_1(\#\Gamma_c(F))) .$$

□

We can thus find the value of $p_1 p_2 q V_P(P_m, Q_n; c)$ by applying the Hungarian or Hopcroft-Karp algorithm to (X', Y', Γ'_c). However, unless p_1 and p_2 are relatively small, the sets X' and Y' will contain so many points that the procedure will be excessively time-consuming, even though the special form of (X', Y', Γ'_c) allows some steps to be saved in applying the algorithms.

The procedure we have described also extends to the case where P (resp. Q) does not give equal probability to the points in X (resp. Y), but the probability of each point is a rational number. In this case there exist positive integers $p_1, \ldots, p_m$ and $q_1, \ldots, q_n$ such that $V_P(P_m, Q_n; c)$ is equal to a constant times the expression

$$\max_{F \subseteq X} \left| \sum_{i=1}^{m} p_i \chi_F(x_i) - \sum_{j=1}^{n} q_j \chi_F(y_j) \right|$$

where χ_F denotes the indicator function of the set F.

In this case we will let X' consist of p_i "copies" of x_i, $i = 1, \ldots, m$, and let Y' consist of q_j "copies" of y_j, $j = 1, \ldots, n$, Γ'_c being defined as before. With the obvious modifications of the proof of 2.1.1, we can show that for this graph (X', Y', Γ'_c),

$$\max_{G \subseteq X'} (\#(G) - \#(\Gamma'_c(G))) = \max_{F \subseteq X} \left(\sum_{i=1}^{m} p_i \chi_F(x_i) - \sum_{j=1}^{n} q_j \chi_{\Gamma_c(F)}(y_j) \right) .$$

In this case, depending on the values of $p_1, \ldots, p_m$ and $q_1, \ldots, q_n$, it may be impractical to find a maximal matching in (X', Y', Γ'_c) even for small values of m and n. In the next section we will treat the special case $S = E^1$ and take advantage of the total ordering on E^1 to develop a more efficient and more generally applicable algorithm.

2.2 Computation of $d_p(P,Q)$ when P has finite support.

Let the support of P be the set $X = \{x_1,\ldots,x_n\}$ where $x_1 \leq x_2 \leq \ldots \leq x_n$. Let Q be an arbitrary measure on E^1. As described before, we wish to find a set $G \subseteq X$ such that $P(G) - Q(G^{c]}) \geq P(F) - Q(F^{c]})$ for all $F \subseteq X$. We shall use an equivalent characterization of such a set G:

Definition 2.2.1: Let $P,Q,$ and c be given. Let $G \subseteq X$ and $F \subseteq X$. If $P(G-F) - P(F-G) \geq Q(G^{c]} - F^{c]}) - Q(F^{c]} - G^{c]})$, we shall say that G is *as bad as* F, and if the strict inequality holds we shall say G is *worse than* F. If G is as bad as every subset of X, we shall call G a *worst set*. □

(Note, if the support of P is infinite, the worst set may not be closed: let $S = E^1$, $c = 1/4$, let P put mass 1/4 at the points -1/4 and 1/4 and be equal to Lebesgue measure on the set (-1/4,1/4). Let Q put mass 1/2 at the points -1/2 and 1/2.) A worst set must maximize $P(G) - Q(G^{c]})$.

Note that if $F \subseteq G \subseteq X$, then G is as bad as F (resp. worse than F) if and only if $Q(G^{c]} - F^{c]}) \leq n^{-1}\#(G-F)$ (resp. $< n^{-1}\#(G-F)$). This will be the criterion of relative badness in most of what follows.

Our algorithm for finding a worst set will examine the points $x_1,\ldots,x_n$ in order to see whether they belong in a worst set. The basic idea is the following.

Proposition 2.2.2: Let i range from 1 to n. Let i' be the least value of i such that either

(a) $Q(\{x_1,\ldots,x_i\}^{c]}) < i/n$, or

(b) $Q(\{x_1,\ldots,x_i\}^{c]} - \{x_{i+1}\}^{c]}) \geq i/n$.

(If $i = n$, replace $\{x_{i+1}\}^{c]}$ by $\emptyset$.) (Note that either (a) or (b) must hold for $i = n$.) If (a) holds for $i = i'$, then x_1 must be included in every worst set; if (b) holds for $i = i'$, then if G is a worst set containing x_1, there is a subset of G which does not contain x_1 and which is also a worst set.

Proof: Assume (a) holds. Let $H \subseteq X$. Let $I = \{x_1,\ldots,x_{i'}\}$. If $I \cap H = \emptyset$, then (a) implies that $I \cap H$ is worse than H. Suppose $I \cap H \neq \emptyset$, but $x_1 \notin H$. Let x_j be the smallest element of $I \cap H$. Since by definition of i', $Q(\{x_1,\ldots,x_{j-1}\}^{c]} - \{x_j\}^{c]}) < (j-1)/n$, $H \cup \{x_1,\ldots,x_{j-1}\}$ is worse than H. Thus $x_1 \notin H$ implies that H is not a worst set.

Assume (b) holds. Let $H \subseteq X$. Suppose $x_1 \in H$. If $I \subseteq H$, then clearly $H - I$ is as bad as H. If $I \nsubseteq H$, then let j' be the greatest value of j, $1 \leq j \leq i'$, such that $x_j \in H$ and

$$\{x_1,\ldots,x_j\}^{c]} \cap [H - \{x_1,\ldots,x_j\}]^{c]} = \emptyset .$$

If no such element exists, let $j' = i'$. Let $J = H - \{x_1,\ldots,x_{j'}\}$.

If $j' = i'$, then $Q(\{x_1,\ldots,x_{j'}\}^{c]} - J^{c]}) \geq Q(\{x_1,\ldots,x_{j'}\}^{c]} - \{x_{j'+1}\}^{c]}) \geq j'/n$, by definition of i', since (b) holds. (This could fail unless $x_1 \ldots x_n$, because we might not have $Q(\{x_1,\ldots,x_{j'}\}^{c]} \geq Q(\{x_1,\ldots,x_{j'}\}^{c]} - \{x_{j'+1}\}^{c]})$.) It follows that J is as bad as H, since $\# J \geq \# H - j'$, while $Q(H^{c]}) - j'/n \geq Q(H^{c]}) - Q(\{x_1,\ldots,x_{j'}\}^{c]} - J^{c]}) \geq Q(J^{c]})$, so $Q(J^{c]}) - Q(H^{c]}) \leq n^{-1}(\# J - \# H)$.

If $j' < i'$, then $Q(\{x_1,\ldots,x_{j}'\}^{c]} - J^{c]}) = Q(\{x_1,\ldots,x_{j}'\}^{c]}) \geq j'/n$, since (a) cannot hold for $j' < i'$, by definition of i'. Again J is as bad as H. □

We next wish to define a sequence of sets $G_0 \subseteq G_1 \subseteq \ldots \subseteq G_n$ such that G_n will be a worst set. Let $G_0 = \emptyset$. Let $G_1 = G_0$ if (b) in 2.2.2 holds for $i = i'$, and let $G_1 = \{x_1\}$ if (a) in 2.2.2 holds for $i = i'$. In general we will have $G_k \subseteq \{x_1,\ldots,x_k\}$.

Assume inductively that G_k is contained in any worst set and that if G is a worst set, then $G - (\{x_1,\ldots,x_k\} - G_k)$ is also a worst set. I.e., G_k satisfies the conditions which 2.2.2 shows are satisfied by $\{x_1\}$. Hence in seeking a worst set, we need only consider points of $G_k \cup \{x_{k+1},\ldots,x_n\}$.

Let i range from 1 to n-k+1. Let i' be the least such value of i such that either

(a') $Q(\{x_{k+1},\ldots,x_{k+i}\}^{c]} - G_k^{c]}) < i/n$, or

(b') $Q(\{x_{k+1},\ldots,x_{k+i}\}^{c]} - G_k^{c]} \cup \{x_{i+1}\}^{c]}) \geq i/n$.

Then if (a') holds for $i = i'$, we let $x_{k+1} \in G_{k+1}$; if (b') holds for $i = i'$, then $x_{k+1} \notin G_{k+1}$. ((a) and (b) are really special cases of (a') and (b') since $G_0 = \emptyset$.)

By the same argument as in 2.2.2 (restricting consideration to sets H such that $\{x_1,\ldots,x_k\} \cap H = G_k$) we can show that G_{k+1} is contained in any worst set and that there always exists a worst set which contains no element of $\{x_1,\ldots,x_{k+1}\} - G_{k+1}$. By induction, G_n will necessarily be a worst set.

Having defined G_k, our algorithm will then be to check for each point x_{k+1} whether (a') or (b') holds for $i = i'$, and to construct the set G_{k+1} accordingly. The algorithm can be improved by noting that if (a') holds so that $x_{k+1} \in G_{k+1}$, then we will have $Q(\{x_{k+2},\ldots,x_{k+i'-1}\}^{c]} - G_{k+1}^{c]}) < (i'-1)/n$ (provided $i' > 1$). Indeed, $i' > 1$ implies that $Q(\{x_{k+1}\}^{c]} - G_k^{c]}) \geq 1/n$. Hence, $x_{k+1} \in G_n$ implies that $x_{k+2},\ldots,x_{k+i'-1}$ are all in G_n, and they need not be tested individually. This improvement could be significant if the evaluation of the measure Q is time-consuming.

The algorithm described in this section is sufficiently efficient to permit the use of Monte Carlo methods to estimate the distribution function of test statistics based on d_P when $S = E^1$. For this purpose, the case of a measure P which puts mass $1/n$ at each of n points is sufficiently general. If instead, P places rational-valued probabilities at each of n points, we can compute $d_P(P,Q)$ by in effect replacing P by a measure which places equal mass on an appropriate number of "copies" of each point of the support -- here "copies" are regarded as distinct observations taking the same value. Unlike the case of general S, for $S = E^1$ this does not require a significant increase in the amount of computation since all the points in each set of "copies" can be tested at the same time; if any is included in the worst set then all must be. (By contrast, in the general case our procedure was to find a maximal matching -- in which only some of a set of copies might appear as initial vertices -- and not to find a worst set.)

2.3 Computation of d_L.

Let $S = E^1$, let $P,Q \in \mathcal{P}(S)$, and let the support of P be the set $X = \{x_1,\dots,x_n\}$ where $x_1 \le x_2 \le \dots \le x_n$. In this case the computation of $d_L(P,Q)$ is simple. By Proposition 1.3.10, $d_L(P,Q) \le c$ if and only if $V_L(P,Q;c) \le c$. Therefore, as with d_P, to be able to find out for each $c \in [0,1]$ whether $V_L(P,Q;c) \le c$, permits computation of $d_L(P,Q)$ to any desired accuracy.

$V_L(P,Q;c)$ is defined as

$$\max\{\sup_{x\in E^1}(Q(-\infty,x-c) - P(-\infty,x]),\ \sup_{x\in E^1}(P(-\infty,x] - Q(-\infty,x+c])\}.$$

As a function of x, $Q(-\infty,x-c) - P(-\infty,x]$ is non-decreasing on the interval $(-\infty,x_1)$ and on each of the intervals $[x_i,x_{i+1})$, $1 \le i \le n-1$. $P(-\infty,x] - Q(-\infty,x+c]$ is non-increasing on each of these intervals. Hence

$$\sup_{-\infty<x<\infty} (Q(-\infty,x-c) - P(-\infty,x]) = \max_{1\le i\le n} (Q(-\infty,x_i-c) - P(-\infty,x_i)) ,$$

and

$$\sup_{-\infty<x<\infty} (P(-\infty,x] - Q(-\infty,x+c]) = \max_{1\le i\le n} (P(-\infty,x_i] - Q(-\infty,x_i+c]) .$$

Thus $V_L(P,Q;c)$ can be computed by evaluating the P-measure of 2n sets and the Q-measure of 2n sets.

If P is not a discrete measure, it may be possible in some special cases of P and Q to determine points x at which

$$\sup_{-\infty<x<\infty} (P(-\infty,x] - Q(-\infty,x+c]) \quad \text{and} \quad \sup_{-\infty<x<\infty} (Q(-\infty,x-c) - P(-\infty,x]) \quad \text{must}$$

occur, provided that a simple functional form for the distribution functions associated with P and Q can be found. Specific examples can readily be suggested, for example, the case where the two density functions are unimodal. However, a useful general characterization of cases in which the maximizing values can be readily determined is not apparent.

If $S = E^k$, we can generalize the method described in this section, although the number of points to be examined increases as n^k, making the method impractical even for moderate values of k. Write the coordinates of x_i as $x_i^{(1)},\ldots,x_i^{(k)}$, for each $i = 1,\ldots,n$. Let $y_1^{(j)} \le y_2^{(j)} \le\ldots\le y_n^{(j)}$ denote the ordered values of the set $\{x_1^{(j)},\ldots,x_n^{(j)}\}$. Let $Z \subseteq E^k$ be the set of all points of the form

$$y = (y_{i_1}^{(1)}\ , y_{i_2}^{(2)},\ldots,y_{i_k}^{(k)})\ ,$$

where $i_1,i_2,\ldots,i_k$ are integers between 1 and n.

$P(-\infty,x] - Q(-\infty,x+c]$ is non-increasing in each coordinate of x in any rectangle of the form

$$[y_{i_1}^{(1)}\ ,y_{i_1+1}^{(1)}\ ) \times [y_{i_2}^{(2)}\ ,y_{i_2+1}^{(2)}\ )\times\ldots\times[y_{i_k}^{(k)},y_{i_k+1}^{(k)})\ ,$$

where $i_1,\ldots,i_k$ are integers between 0 and n, y_0 being given the meaning of replacing the left-hand endpoint by $-\infty$. Hence

$$\sup_{x\in E^k}(P(-\infty,x] - Q(-\infty,x+c]) = \max_{y\in Z}(P(-\infty,y] - Q(-\infty,y+c])\ .$$

Likewise, $Q(-\infty,x-c) - P(-\infty,x]$ is a non-decreasing function of each coordinate of x in each of the rectangles described and hence

$$\sup_{x\in E^k}(Q(-\infty,x-c) - P(-\infty,x]) = \max_{y\in Z}(Q(-\infty,y-c) - P(-\infty,y))\ .$$

As before, the problem is reduced to evaluating the P-measure and Q-measure of a finite number of sets. $\#(Z)$ can be as large as n^k.

If Q has finite support, the problem can be treated by the methods of Section 2.1, dealing separately with the two quantities

(a) $\inf\{\varepsilon > 0: P(-\infty,x] - Q(-\infty,x+c] \le c;\ \text{for all}\ x \in E^k\}$

and

(b) $\inf\{\varepsilon > 0: Q(-\infty,x-c) - P(-\infty,x] \le c;\ \text{for all}\ x \in E^k\}$.

Let $X = \{x_1,\ldots,x_n\}$ be the support of P and let $Y = \{y_1,\ldots,y_n\}$ be the support of Q. We then define a graph (X,Y,Γ_c) by $(x_i,y_j) \in \Gamma_c$ if and only if $y_j \le x_i + c$, so that $nD(X,Y,\Gamma_c) = \max_{F\subseteq X}(P(F) - Q(\Gamma_c(F)))$ $= \max_{x\in Z}(P(-\infty,x] - Q(-\infty,x+c])$. This deals with (a). For (b), let the arc $(y_j,x_i) \in \Gamma_c$ if and only if $y_j + c < x_i$, and again $nD(Y,X,\Gamma_c) =$ $\max_{F\subseteq Y}(Q(F) - P(\Gamma_c(F))) = \max_{x\in Z}(Q(-\infty,x-c) - P(-\infty,x))$. In each case, the deficiency may be found by the Hungarian or the Hopcroft-Karp algorithm.

The extension to the cases of unequal numbers of atoms for P and Q, or for atoms not all the same size but rational-valued, is done in the same way as in Section 2.1.

CHAPTER 3

Null Distribution for the Two-Sample Test Based on d_P

3.1 Introduction.

As described in Section 1.1, the null distribution of our two-sample test statistic based on d_P is, for equal sample sizes, the distribution of $d_P(\mu_n, \nu_n)$ where μ_n and ν_n are the empirical measures of $r_1/2n, \ldots, r_n/2n$ and $s_1/2n, \ldots, s_n/2n$, $r_1 < \ldots < r_n$ being a random sample without replacement from $\{1,2,\ldots,2n\}$ and $s_1 < \ldots < s_n$ being the elements in order of $\{1,2,\ldots,2n\} - \{r_1, \ldots, r_n\}$. For each $c \in [0,1]$, we define the graph (X,Y,Γ_c) by letting $X = \{r_1/2n, \ldots, r_n/2n\}$, $Y = \{s_1/2n, \ldots, s_n/2n\}$ and $(r_i/2n, s_j/2n) \in \Gamma_c$ if and only if $|r_i/2n - s_j/2n| \le c$, i.e., if and only if $|r_i - s_j| \le i.p.\ (2nc)$, where "*i.p.*" denotes "integer part of." The following can be stated immediately based on the results of sections 1.3 and 1.5.

Proposition 3.1.1: For positive integers p, n, and z, let $A(p,n,z)$ be the number of distinct samples $\{r_1, \ldots, r_n\}$ and $\{s_1, \ldots, s_n\}$ such that $D(X,Y,\Gamma_{z/2n}) = p$. Then for $z/2n \le c < (z+1)/2n$,

$$\Pr\{d_P(\mu_n, \nu_n) \le c\} = \Pr\{D(X,Y,\Gamma_{z/2n}) \le z/2\} = \binom{2n}{n}^{-1} \sum_{0 \le p \le z/2} A(p,n,z) .$$

□

Proposition 3.1.1 can be generalized to the case of samples which are not real-valued. However, for $S = E^1$, the structure of Γ_c is sufficiently simple to permit the evaluation of $A(p,n,z)$; the general case seems unmanageable, as does the case of unequal sample sizes. In the remainder of this chapter, we will discuss the features of Γ_c which permit this evaluation to be performed using the simple relationship expressed in Proposition 3.1.2 below.

Let X be a countable set such that each element x of X has a unique classification $\kappa(x)$ in a "classification set" C. We wish to determine for each $u \in C$ the number $K(u)$ of elements of X which are classified as u, i.e., $K(u)$ is defined as $\#(\kappa^{-1}(u))$. To this end we predicate the existence of a mapping $a\colon X \to X$ having the property that for each $u,v \in C$, $x,y \in \kappa^{-1}(u)$ implies that $\#\{a^{-1}(x) \cap \kappa^{-1}(v)\} = \#\{a^{-1}(y) \cap \kappa^{-1}(v)\}$. Thus for each $x \in \kappa^{-1}(u)$ the number of elements $z \in X$ for which $a(z) = x$ and $\kappa(z) = v$ depends only on u and v, not on the choice of x. Call this number $T(u,v)$. (If $K(u) = 0$, let $T(u,v) = 0$).

<u>Proposition 3.1.2</u>: $K(v) = \sum_{u \in C} K(u)T(u,v)$, for each $v \in C$.

<u>Proof</u>: Observe that $K(u)T(u,v)$ is the number of elements $w \in X$ such that $\kappa(w) = v$ and $\kappa(a(w)) = u$. But $\kappa^{-1}(v) = \sum_{u \in C} \{w\colon \kappa(w) = v \text{ and } \kappa(a(w)) = u\}$. □

For a fixed z, our problem of computing $A(p,n,z)$ for various values of n and p can now be expressed in terms of Proposition 3.1.2. X will be the set $\bigcup_{n=0}^{\infty} \{(X,Y)\colon (X,Y)$ is a partition of $\{1,2,\ldots,2n\}$ into subsets X and Y with $\#(X) = \#(Y) = n\}$. The case $n = 0$ we take to represent a partition of the empty set, which we will identify with the empty set. Note that the sets X and Y are now sets of integers, not multiples of $1/2n$.

Each element u of the classification set will consist of a 5-tuple of integers $u = (k,n,d,s,t) = (k(u),n(u),d(u),s(u),t(u))$, describing various properties which hold for each partition (X,Y) classified as u. In particular n has the meaning of the previous paragraph. $d(u)$ is the deficiency of the graph (X,Y,Γ_z). Γ_z is similar to $\Gamma_{z/2n}$ and will be defined below, as will the parameters s and t. The mapping $a: X \to X$ will be defined in such a way that $T(u,v) = 0$ whenever $n(u) \geq n(v)$, except when $u = \kappa(\phi)$, in which case $T(u,u) = 1$, i.e., we define $a(\phi) = \phi$. Only for the classification $u = \kappa(\phi)$ is $n(u) = 0$; we define $K(\kappa(\phi)) = 1$. Hence, if $T(u,v)$ can be computed for all u and v in C, then $K(u)$ can be found for all $u \in C$. We then have

$$A(p,n,z) = \sum_k \sum_s \sum_t K(k,n,p,s,t) .$$

In the next section we define a classification set and a mapping a in such a way that $T(u,v)$ can be computed for all u and v.

3.2 Matchings in the graphs (X,Y,Γ_z).

Our ultimate interest is in partitions (X,Y) of $\{1,2,\ldots,2n\}$ into sets X and Y of equal size. For use in later proofs we wish to describe properties which apply more generally to partitions of $\{1,2,\ldots,n+n'\}$ into sets X and Y with $\#(X) = n$ and $\#(Y) = n'$.

The positive integer z will be considered to be fixed and we will hereafter use the convention that for any partition (X,Y) under consideration, Γ_z will denote the graph such that $(x,y) \in \Gamma_z$ whenever $|x-y| \leq z$ for $x \in X$ and $y \in Y$. We write $\Gamma_z(X,Y)$ if the partition is not clear from the context.

We shall deal with $D(X,Y,\Gamma_z)$ by looking at maximal matchings in (X,Y,Γ_z), using Proposition 1.5.1. Since we are really interested only in $D(X,Y,\Gamma_z)$, rather than the matchings, it will be convenient to consider a single "canonical" maximal matching, uniquely defined for each partition (X,Y), and having certain special properties, which we will describe in Proposition 3.2.14 following some discussion.

Definition 3.2.1: Let (X,Y) partition $\{1,2,\ldots,n+n'\}$ so that $\#(X) = n$, $\#(Y) = n'$, and let $M = N(X,Y,\Gamma_z)$. Let $m \in \Gamma_z$ be a maximal matching and write m as the set of arcs $\{(x_1(m),y_1(m)),\ldots,(x_M(m),y_M(m))\}$, such that $x_1(m) < x_2(m)<\ldots<x_M(m)$. (The elements $y_1(m),\ldots,y_M(m)$ are not necessarily in order.) Define $X^+(m) = \{x_1(m),\ldots,x_M(m)\}$, $Y^+(m) = \{y_1(m),\ldots,y_M(m)\}$, $X^-(m) = X-X^+(m)$, and $Y^-(m) = Y-Y^+(m)$. □

We will be chiefly interested in the sets X^+ and Y^+ and not in the particular matching which produced them. In fact, we will always be able to assume that the first element of X^+ is connected to the first element of Y^+, second to second, etc.

Proposition 3.2.2: Let m be a maximal matching in (X,Y,Γ_z). Then there exists a maximal matching $m^* \subseteq \Gamma_z$ such that $X^+(m) = X^+(m^*)$, $Y^+(m) = Y^+(m^*)$, and $y_1(m^*) < y_2(m^*)<\ldots<y_M(m^*)$.

Proof: Follows from Lemma 3.2.3.

Lemma 3.2.3: Let $y_{i'}$ and $y_{j'} \in Y$ and $x_i,x_j \in X$ be such that $(x_i,y_{i'}) \in \Gamma_z$ and $(x_j,y_{j'}) \in \Gamma_z$. If $x_i \le x_j$ and $y_{j'} \le y_{i'}$, then

$(x_i, y_{j'}) \in \Gamma_z$ and $(x_j, y_{i'}) \in \Gamma_z$. (The roles of X and Y can be interchanged.)

<u>Proof</u>: It follows from the inequalities $|x_i - y_{i'}| \le z$ and $|x_j - y_{j'}| \le z$ and the assumption $x_i \le x_j$, $y_{j'} \le y_{i'}$, that $|x_i - y_{j'}| \le z$ and $|x_j - y_{i'}| \le z$. □

<u>Definition 3.2.4</u>: Let $\mathcal{M}(X,Y,\Gamma_z)$ be the collection of all maximal matchings $m \subseteq \Gamma_z$ such that $y_1(m) < \dots < y_M(m)$. □

A matching in $\mathcal{M}$ is uniquely determined by its sets $X^+(m)$ and $Y^+(m)$ and vice versa. However, for a given partition (X,Y), there may be a number of matchings in $\mathcal{M}(X,Y,\Gamma_z)$.

We will introduce a schematic notation to be used in describing the partitions and matchings under consideration and use it to give some simple examples showing considerations involved in choosing a "canonical" matching. yxy will denote a partition of $\{1,2,3\}$ into the sets $X = \{2\}$ and $Y = \{1,3\}$. Here $n = 1$ and $n' = 2$. For $z \ge 1$, we have two matchings, denoted by $\overset{+}{y}\overset{+}{x}\overset{-}{y}$ and $\overset{-}{y}\overset{+}{x}\overset{+}{y}$, by which we mean respectively $\{(2,1)\}$ and $\{(2,3)\}$. In using our notation we always assume that the elements of X^+ and Y^+ are connected in order, i.e., the matching is in $\mathcal{M}$. Of the two matchings given, $\overset{+}{y}\overset{+}{x}\overset{-}{y}$ will be chosen as the "canonical" one, because it is "earliest" (the +'s occur earlier), in a sense to be made precise. From among $\overset{+}{y}\overset{-}{y}\overset{+}{x}\overset{-}{y}$, $\overset{-}{y}\overset{+}{y}\overset{+}{x}\overset{-}{y}$, and $\overset{-}{y}\overset{-}{y}\overset{+}{x}\overset{+}{y}$, the matching $\overset{+}{y}\overset{-}{y}\overset{+}{x}\overset{-}{y}$ is the "earliest." However, in $\overset{-}{y}\overset{+}{y}\overset{+}{x}\overset{-}{y}$ and $\overset{-}{y}\overset{-}{y}\overset{+}{x}\overset{+}{y}$ the matched x and y are as close together as possible,, while in $\overset{+}{y}\overset{-}{y}\overset{+}{x}\overset{-}{y}$, there is an intervening element of Y^-. $\overset{+}{y}\overset{-}{y}\overset{+}{x}\overset{-}{y}$ will be said not to be "compressed," while the other two matchings will be said to be "compressed," a notion to be made precise below. In this case, the "canonical" matching will be chosen to be the "earlier" of the two "compressed" matchings, namely $\overset{-++-}{yyxy}$. It will later be seen that "compressed" matchings assign +'s and -'s in a useful pattern. The choice of the "earliest" of the "compressed" matchings is made for definiteness.

Corresponding to the notion of earliest matching, we now introduce a total ordering on $M(X,Y,\Gamma_z)$ which we shall call the *lexicographic ordering*. In this ordering we say that $m < m'$ if the vector $(x_1(m),\ldots,x_M(m),y_1(m),\ldots,y_M(m))$ is *lexicographically smaller* than $(x_1(m'),\ldots,x_M(m'),y_1(m'),\ldots,y_M(m'))$, i.e., if $x_1(m) < x_1(m')$, or $x_1(m) = x_1(m')$ but $x_2(m) < x_2(m')$, or $x_1(m) = x_1(m')$ and $x_2(m) = x_2(m')$ but $x_3(m) < x_3(m')$, etc.

<u>Proposition 3.2.5</u>: Let m be the least element of $M(X,Y,\Gamma_z)$ with respect to the ordering just defined and let m' be any element of M, Then $x_1(m) \le x_1(m'),\ldots,x_M(m) \le x_M(m')$, $y_1(m) \le y_1(m'),\ldots,y_M(m) \le y_M(m')$. (Note that the definition of lexicographic minimality only implies that $x_1(m) \le x_1(m')$. If $m \notin M$ the result may not hold, for example if m is $\overset{+++-+}{yxxxy}$, m' could be $\{(3,5),(4,1)\}$ provided $z \ge 3$.)

<u>Proof</u>: Necessarily $x_1(m) \le x_1(m')$. Further, $y_1(m) \le y_1(m')$. Indeed, if $y_1(m) > y_1(m')$, then by Lemma 3.2.3, we can in $Y^+(m)$ replace $y_1(m)$ by $y_1(m')$ (since $y_1(m') \in Y^-(m)$) and get a matching in M which is lexicographically smaller than m.

Let j be the least index, $2 \le j \le M$, such that $x_j(m) > x_j(m')$, or $y_j(m) > y_j(m')$. We wish to arrive at a contradiction by assuming the existence of such a j. Suppose first that $x_j(m) > x_j(m')$. Since j

is the first index of its kind, $x_{j-1}(m) \le x_{j-1}(m') < x_j(m') < x_j(m)$. Hence $x_j(m') \in X^-(m)$. Similarly, $y_{j-1}(m') \le y_{j-1}(m) < y_j(m)$.

We have two cases:

a) $y_j(m) > y_j(m')$;

b) $y_j(m) \le y_j(m')$.

In case (b), we can apply Lemma 3.2.3 to show that $(x_j(m'), y_j(m)) \in \Gamma_Z$, because $x_j(m) > x_j(m')$ and $y_j(m) \le y_j(m')$. Hence we can in $X^+(m)$ replace $x_j(m)$ by $x_j(m')$, which is in $X^-(m)$, to get a lexicographically smaller matching in M , contradicting the definition of m. In case (a), we contradict the maximality of m, since in this case we must have $y_j(m') \in Y^-(m)$, as was shown above for $x_j(m')$. Since either (a) or (b) leads to a contradiction, we must have $x_j(m) \le x_j(m')$.

So assume that $x_j(m) \le x_j(m')$, but $y_j(m) > y_j(m')$. Then as before $y_j(m') \in Y^-(m)$. By Lemma 3.2.3 we may replace the element $y_j(m)$ in $Y^+(m)$ by $y_j(m')$, obtaining an earlier matching than m, contradicting the definition of m. □

We now define the class of *compressed* matchings.

<u>Definition 3.2.6</u>: Let $M'(X,Y,\Gamma_Z)$ be the class of all $m \in (X,Y,\Gamma_Z)$ such that $X^+(m) + Y^+(m) = \bigcup_{i=1}^{\infty} [\min(x_i(m),y_i(m)), \max(x_i(m),y_i(m))]$, where $[i,j]$ denotes the integers between i and j inclusive. (I.e., none of the intervals $[\min(x_i(m),y_i(m)), \max(x_i(m),y_i(m))]$ contains any point of $X^-(m) + Y^-(m)$.) □

The lexicographic ordering on M induces an ordering on M' as a subset of M. We wish to show that the least element m of M' in this ordering satisfies the stronger conditions of Proposition 3.2.5 when compared to any other element of M'. To do this we first give an algorithm which associates with any element $m \in M$ a unique element m^c of M'.

Definition 3.2.7: Let $m \in M(X,Y,\Gamma_z)$. We define the element $m^c \in M'(X,Y,\Gamma_z)$ by the following algorithm.

Let $m^{(0)} = m$. Suppose $m^{(0)},\ldots,m^{(i)}$ have been defined. If $m^{(i)} \in M'$, let $m^c = m^{(i)}$; we have completed the algorithm. If $m^{(i)} \notin M'$, we define $m^{(i+1)}$ as follows. Let j be the least index, $1 \le j \le M$, such that $[\min(x_j(m^{(i)}), y_j(m^{(i)})), \max(x_j(m^{(i)}), y_j(m^{(i)}))] \not\subseteq X^+(m^{(i)}) + Y^+(m^{(i)})$. We have one of four mutually exclusive cases:

a) there exists an $x \in X^-(m^{(i)})$ such that $x_j(m^{(i)}) < x < y_j(m^{(i)})$

b) there exists an $x \in X^-(m^{(i)})$ such that $y_j(m^{(i)}) < x < x_j(m^{(i)})$

c) there exists a $y \in Y^-(m^{(i)})$ such that $x_j(m^{(i)}) < y < y_j(m^{(i)})$

d) there exists a $y \in Y^-(m^{(i)})$ such that $y_j(m^{(i)}) < y < x_j(m^{(i)})$.

The cases are mutually exclusive because if, for example, (a) and (c) both hold, then $(x,y) \in \Gamma_z$, contradicting the maximality of m.

x (or y) need not be unique. Let $x°$ (resp. $y°$) be the largest such x (resp. y) in case (b) (resp. (c)) and the smallest such x (resp. y) in case (a) (resp. (d)). Define $m^{(i+1)}$ in cases (a) and (b) (resp. (c) and (d)) by replacing, in $X^-(m^{(i)})$ (resp. $Y^-(m^{(i)})$) , $x_j(m^{(i)})$ by $x°$ (resp. $y_j(m^{(i)})$ by $y°$). □

We must show that the procedure described in 3.2.7 eventually produces an element of M'. To do this, define the function g on M

by

$$g(m) = \sum_{i=1}^{M} |x_i(m)-y_i(m)| .$$

We claim that (unless $m^{(i)} \in M'$)

$$g(m^{(i)}) > g(m^{(i+1)}) ;$$

it will follow that since the function g is bounded below by 0, the algorithm will continue until g can no longer be decreased.

Our claim is an immediate consequence of the following result.

Lemma 3.2.8: (i) In case (a) above, let $l \geq 0$ be the integer such that

$$x_1(m^{(i)})<\ldots<x_j(m^{(i)})<\ldots<x_{j+l}(m^{(i)}) < x < x_{j+l+1}(m^{(i)})<\ldots<x_M(m^{(i)}) .$$

Then $m^{(i+1)}$ consists of the pairs $(x_1(m^{(i)}), y_1(m^{(i)})),\ldots,$ $(x_{j-1}(m^{(i)}),y_{j-1}(m^{(i)}))$ and if $l > 0$, $(x_{j+1}(m^{(i)}),y_j(m^{(i)})),\ldots,$ $(x_{j+l}(m^{(i)}), y_{j+l-1}(m^{(i)}))$, and (for $l \geq 0$), $(x^\circ,y_{j+l}(m^{(i)}))$ and $(x_{j+l+1}(m^{(i)}),y_{j+l+1}(m^{(i)})),\ldots,(x_M(m^{(i)}),y_M(m^{(i)}))$.

(ii) In case (b) above, $m^{(i+1)}$ is identical to $m^{(i)}$ except that the arc $(x_j(m^{(i)}),y_j(m^{(i)}))$ is replaced by $(x^\circ,y_j(m^{(i)}))$.

The cases (c) and (d) interchange the role of X and Y but are otherwise identical.

Proof: (i) is obvious when one follows through Definition 3.2.7.

(ii) Because j is the least integer such that there is an element $x \in X^-(m^{(i)})$ such that $y_j(m^{(i)}) < x < x_j(m^{(i)})$, we cannot have $x < x_{j-1}(m^{(i)}) < x_j(m^{(i)})$, because then $y_{j-1}(m^{(i)}) < x < x_{j-1}(m^{(i)})$.
The result is then obvious. □

Proposition 3.2.9: Let $m^\circ \in M'(X,Y,\Gamma_z)$ be the lexicographically minimal element of M'. Let $m' \in M'$ be any other element. Then

$$x_1(m^\circ) \le x_1(m'),\ldots,x_M(m^\circ) \le x_M(m'),y_1(m^\circ) \le y_1(m'),\ldots,y_M(m^\circ) \le y_M(m').$$

Proof: The result will be an immediate corollary of the following.

Lemma 3.2.10: Let $m' \in M'$ and $m \in M$ such that $x_i(m) \le x_i(m')$ and $y_i(m) \le y_i(m')$, $i=1,\ldots,M$. Then $x_i(m^c) \le x_i(m')$ and $y_i(m^c) \le y_i(m')$, $i=1,\ldots,M$.

Proof: It suffices to prove the conclusion for $m^{(1)}$ instead of m^c, since $m^{(2)} = (m^{(1)})^{(1)}$, etc. We prove this result for $m^{(1)}$ in the cases (a) and (b) above, the cases (c) and (d) following by interchanging the roles of X and Y.

Suppose (a) obtains. We distinguish two further cases:

A) $x_j(m') < y_j(m')$

B) $x_j(m') > y_j(m')$.

Suppose (A) holds. Let l be as in 3.2.8(a). Since x° is the smallest element of $X^-(m)$ between $x_j(m)$ and $y_j(m)$, there are $l+1$ elements of X in the half-open interval $(x_j(m),x^\circ]$, namely $x_{j+1}(m),\ldots,x_{j+l}(m)$, and x°. Since $x^\circ < y_j(m) \le y_j(m')$, and because $m' \in M'(X,Y,\Gamma_z)$, if $x_j(m') < x^\circ$, it follows that all elements of X in the interval $(x_j(m'),x^\circ]$ must be elements of $X^+(m')$. But because $x_{p+l+1}(m') \ge x_{p+l+1}(m) > x^\circ$, we have at most l elements of $X^+(m')$ and hence of X in the interval $(x_j(m'),x^\circ]$. Hence, if $x_j(m') < x^\circ$, then $x_j(m) < x_j(m')$. Of course, if $x_j(m') \ge x^\circ$, then $x_j(m) < x_j(m')$.

In either case, then, $x_{j+1}(m) \le x_j(m'),\ldots,x_{j+l}(m) \le x_{j+l-1}(m')$, and $x^\circ \le x_{j+l}(m')$. Therefore, in making the replacements of arcs described in 3.2.8(i), we have

$$x_j(m^{(1)}) \le x_j(m'),\ldots,x_{j+l}(m^{(1)}) \le x_{j+l}(m') ,$$

and

$$Y^+(m) = Y^+(m^{(1)}) .$$

Suppose (B) holds. Since $x^\circ < y_j(m) \le y_j(m') < x_j(m')$, we again have

$$x_j(m^{(1)}) \le x_j(m'),\ldots,x_{j+l}(m^{(1)}) \le x_{j+l}(m').$$

Suppose that case (b) obtains. Only one arc, (x_j,y_j) is changed in going from m to $m^{(1)}$. But $x_j(m^{(1)}) \le x_j(m)$ and $y_j(m^{(1)}) = y_j(m)$, so $x_j(m^{(1)}) \le x_j(m')$ and $y_j(m^{(1)}) = y_j(m')$. □

Proposition 3.2.11: $m^\circ = m^c$, where m is the minimal matching in $M(X,Y,\Gamma_z)$.

Proof: Follows immediately from 3.2.10. □

The matching m° will be our canonical matching for the partition (X,Y). For each (X,Y), the sets $X^+(m^\circ)$, $Y^+(m^\circ)$, $X^-(m^\circ)$, $Y^-(m^\circ)$ are uniquely determined. The choice of m° as canonical matching establishes a certain pattern in the arrangement of the elements of these sets. For example, let $z = 2$ and consider $\overset{----++++-++++---}{xxxxxxyyyyxyxyyy}$. The elements of $X^+ + Y^+$ are arranged in two segments of consecutive elements, namely {5,6,7,8} and {10,11,12,13}. In general, such segments must each contain as many elements of X^+ as of Y^+, since otherwise an element of $X^- + Y^-$ would lie in one of the intervals $[\min(x_i(m^\circ),y_i(m^\circ)), \max(x_i(m^\circ),y_i(m^\circ))]$. The segments are separated by one or more elements of X^- or Y^-, but clearly an element of X^- cannot be adjacent to an element of Y^-. Other remarks can be made, but we wish to proceed more formally and collect the results in Proposition 3.2.14.

Definition 3.2.12: For each partition (X,Y) of {1,2,...,n+n'} define the function $f = f_{(X,Y)}$ on {0,1,...,n+n'} by $f(0) = 0$ and

$$f(j) = \#(X\cap\{1,2,\ldots,j\}) - \#(Y\cap\{1,2,\ldots,j\}), \quad \text{for } j=1,2,\ldots,n+n'.$$

Let I be the interval of integers $[j_0+1,\ j_0+l+l']$ consisting of $l+l'$ consecutive integers from the set {1,2,...,n+n'}, l of X and l' of Y, where $0 \le j_0 \le n+n' - 1$, $l \ge 0$ and $l' \ge 0$. I will be said to be *non-switching* if either

$$f(j) - f(j_0) \ge 0, \quad \text{for all } j \in [j_0+1,\ j_0+l+l'] \quad \text{or if}$$
$$f(j) - f(j_0) \le 0, \quad \text{for all } j \in [j_0+1,\ j_0+l+l'] .$$

The interval I will be said to be *strictly non-switching* if the strict inequality holds for all $j \in [j_0+1,\ j_0+l+l')$. □

Examples: switching xxyyyx

non-switching (not strict) xxyyxy

non-switching (strict) xxxyyy .

Note that if $[j_0+1, j_0+2l]$ is a non-switching interval containing l elements of X and l elements of Y, then $j_0+1 \in X$ if and only if $j_0+2l \in Y$.

Definition 3.2.13: Let (X,Y) partition {1,2,...,n+n'} as before. The elements 1,2,...,n+n' will be divided into the sets $A_1 < B_1 < A_2 < \ldots < A_r < B_r < A_{r+1}$, for some value of r, where each set $A_i \subseteq X^- + Y^-$ and each set $B_i \subseteq X^+ + Y^+$. Let q_1 be the least element of $X^+(m°) + Y^+(m°)$. Let B_1 be the largest non-switching interval of

$X^+(m°) + Y^+(m°)$ which has q_1 as its left-hand endpoint. Let q_2 be the first element (if one exists) of $X^+(m°) + Y^+(m°)$ which is greater than the right-hand endpoint of B_1 and let B_2 be the largest non-switching interval of $X^+(m°) + Y^+(m°)$ which has q_2 as its left-hand endpoint. Let q_i and B_i be defined in this manner until the union $B_1 + B_2 + \ldots + B_r$ contains all the elements of $X^+(m°) + Y^+(m°)$.

Let A_1 be defined as the interval $[1,q_1)$ and let A_i, for $i=2,3,\ldots,r$, be the interval of elements of $X^- + Y^-$ which lie between the right-hand endpoint of B_{i-1} and the left-hand endpoint of B_i, and let A_{r+1} be the interval of elements to the right of B_r. Any or all of the sets $A_1,A_2,\ldots,A_{r+1}$ may be empty. □

Proposition 3.2.14:

(a) For $i=1,\ldots,r$, B_i consists of the same number of elements of X as of Y.

(b) For $i=2,\ldots,r$, if $A_i = \phi$, then $q_i \in X$ if and only if $q_{i-1} \in Y$.

(c) Either $A_i \subseteq X^-(m°)$ or $A_i \subseteq Y^-(m°)$ for $i=1,2,\ldots,r+1$.

(d) If $A_i \neq \phi$, then $q_i \in X$ if and only if $A_i \subseteq X$, for $i=1,2,\ldots,r$.

Proof: (a) and (c) are as noted above. (b) holds because otherwise B_{i-1} would not be the largest non-switching interval starting at q_{i-1}. For (d) we need a preliminary Lemma.

Definition 3.2.15: Let the interval $I = [j_0+1, j_0+2l]$ consist of l elements of X and l of Y. I will be said to be *completely matchable* under Γ_z if the graph $(I \cap X, I \cap Y, \Gamma_z)$ has zero deficiency.

□

<u>Lemma 3.2.16</u>: Let I be a strictly non-switching interval which is completely matchable under Γ_z. Then $[j_0+2,\ j_0+2l-1]$ is completely matchable under Γ_z.

<u>Proof</u>: Assume for definiteness that $j_0+1 \in X$. Let f be as in 3.2.9. Let $a_1 < a_2 < \ldots < a_l$ be the elements of $I \cap X$ and $b_1 < b_2 < \ldots < b_l$ be the elements of $I \cap Y$. Then for $i=2,\ldots,l$, $a_i < b_{i-1}$; otherwise $f(b_{i-1}) - f(j_0) \le 0$ and I would not be strictly non-switching. But for $i=2,\ldots,l$, $|a_i - b_i| \le z$, since I is completely matchable under Γ_z. Hence, since $a_i < b_{i-1} < b_i$, $(a_i, b_{i-1}) \in \Gamma_z$. The matching $\{(a_{i+1}, b_i),\ 1 \le i \le l-1\}$ is a complete matching on $[j_0+2, j_0+2l-1]$.

□

<u>Continuation of 3.2.14</u>: Assume that (d) does not hold for some i and for definiteness that $q_i \in X$ and $A_i \subseteq Y$. Let J be the largest strictly non-switching interval starting at q_i. Then $J \subseteq B_i$.

We can form a matching m which is the same as m° except that q_i is matched to the last element of A_i, the last element of J is unmatched, and the remaining elements of J are matched to each other as in Lemma 3.2.16. Then $m \in M'(X,Y,\Gamma_z)$, but m is lexicographically smaller than m°, contradicting the definition of m°. □

(As an illustration of this last point, take $r=2$, $i=1$ and $z=3$, and consider $\overset{---++++++}{yyyxxyyyx}$. The matching shown is not m° because $\overset{--++++-++}{yyyxxyyyx}$ is a lexicographically smaller compressed matching.)

3.3 Definition of a mapping $a: \mathcal{X} \to \mathcal{X}$.

Let $\mathcal{X}_n = \{(X,Y) : (X,Y)$ is a partition of $\{1,2,\ldots,2n\}$ such that $\#(X) = \#(Y) = n\}$. Then $\mathcal{X} = \bigcup_{n=0}^{\infty} \mathcal{X}_n$, where $\mathcal{X}_0 = \{\phi\}$. Let $x = (X,Y) \in \mathcal{X}_n$ and let $j = \#(X \cap B_r(X,Y,\Gamma_z)) = \#(Y \cap B_r(X,Y,\Gamma_z))$, where $B_r(X,Y,\Gamma_z)$ is the set B_r defined in Definition 3.2.13. Then $a(x)$ will be that element of $\mathcal{X}_{n-j}$ which is the same as (X,Y) except that B_r has been "removed" and the elements of A_{r+1} have been "pushed forward" and "relabeled" in the obvious manner. (In our schematic notation if $z = 2$ and x is $\overset{----++++--++++++--}{\mathrm{xxxxxxyyyyyx\underline{xxyy}yy}}$ with B_r underlined, then $a(x)$ will be the element xxxxxxyyyyyxyy.)

In describing the relationship of (X,Y) to $a(X,Y)$ it will be convenient to adopt some suggestive terminology. $a(X,Y)$ will be said to have been obtained by *removing* B_r from (X,Y). Conversely, (X,Y) will be said to have been obtained by *inserting* B_r (or really a copy of B_r) *into* $a(X,Y)$ *after the integer* $q_r(X,Y,\Gamma_z) - 1$. For purposes of proof we shall wish to consider in the same sense "removals" and "insertions" of arbitrary numbers of X-type and Y-type elements, not just completely matchable intervals.

We now return to Proposition 3.1.2. $a^{-1}(X,Y)$ will consist precisely of every element (X',Y') of $\mathcal{X}$ which we can get by inserting an interval into (X,Y) such that the inserted interval coincides exactly with $B_r(X',Y',\Gamma_z)$. Our interest will be in sets of elements of $\mathcal{X}$ of the form $a^{-1}(X,Y) \cap \kappa^{-1}(v)$, in particular in how the deficiency of each element of $a^{-1}(X,Y)$, which is given by one of the parameters of its classification v, relates to the deficiency of elements of $\kappa^{-1}(\kappa(X,Y))$, all of which have the same deficiency.

There is one further complication. Consider the inverse procedure of going from a(X',Y') to (X',Y') by insertion of an interval. Depending on the location of the insert, it may be the case that not just any interval can be inserted so as to coincide with B_r(X',Y'). For example, we may insert xxyy into (X,Y) = xxxyyxyy after 6 to get (for z = 2) (X',Y') = $\overset{-++++-++++--}{\text{xxxyyx}\underline{\text{xxyy}}\text{yy}}$, and have a(X',Y') = (X,Y). But if (for z = 2) we insert xy after 6 to get $\overset{-++++++++-}{\text{xxxyyx}\underline{\text{xy}}\text{yy}}$, we then have a(X',Y') = xxxyyy , which is not the original element (X,Y), i.e., $B_r(X',Y',\Gamma_z)$ does not coincide with the inserted interval.

In this example, xxyy is intuitively "long enough" to separate element 6 (in X) from element 7 (in Y) so that after the insertion these elements cannot be part of the same matched segment, while xy is not long enough. However, length of the inserted interval is not a sufficient condition for the interval to be inserted in such a location.

<u>Definition 3.3.1</u>: A partition (X,Y) of $\{1,2,\ldots,2l\}$ with #(X) = #(Y) will be said to be *separating* if

a) (X,Y) is completely matchable under Γ_z .

b) The partition (X',Y') obtained by inserting (X,Y) after 1 into xy (assuming for definiteness that $1 \in X$; if $1 \in Y$, then insert into yx) is not completely matchable under Γ_z. □

Thus for z = 2, the interval xxyy is separating because $\overset{-++++-}{\text{xxxyyy}}$ is not completely matchable. xy is not separating because xxyy is completely matchable. (Note that for z = 3, xxxyyy (and also xxyxyy) are separating but xyxyxy is not, although it is of the same length.)

3.4 Classification scheme for elements of $\mathcal{X}$.

Suppose that

a) $(X,Y) \in \mathcal{X}_n$,

b) (X,Y,Γ_z) has deficiency d,

c) $B_r(X,Y,\Gamma_z)$ ends with s consecutive elements of either X or Y, $s \geq 1$,

d) $\#A_{r+1}(X,Y,\Gamma_z) = t$.

Then $\kappa(X,Y) = u = (k,n,d,s,t)$, where $\mathcal{X}$ is divided into three subsets corresponding to $k = 0,1,$ or 2 according to which of the following conditions is satisfied.

A) For some i^*, $1 \leq i^* \leq r(X,Y,\Gamma_z)$, $A_{i^*} \neq \phi$ and $A_{i^*+1} = \phi,\ldots,$ $A_r = \phi$ and $B_{i^*}(X,Y,\Gamma_z)$ is a separating interval.

B) $A_{r+1} = \phi$, or $A_{r+1} \neq \phi$ and $A_{r+1} \subseteq X$ if and only if the final element of B_r is classified as X.

C) $A_{r+1} \neq \phi$ and $A_{r+1} \subseteq X$ if and only if the final element of B_r is classified as Y.

D) $A_i = \phi$ for all i, $1 \leq i \leq r$.

$k = 0$ when (D) holds or (A) and (B) hold.

$k = 1$ when (A) and (C) hold.

$k = 2$ when (C) holds, but (A) does not hold.

Examples: (B_r is underlined) $z = 3$.

```
         ----+++++++-+++++++---
k = 0:   yyyyyyyxxxxxxxyyyyxxxx     (A) and (B) hold.  i* = 2.

         -++++++----++++++++---
k = 1:   yyyyxxxxxxxxxxyyyyxyyy     (A) and (C) hold.  i* = 2.

         ----++++++--++++---
k = 2:   yyyyyyyxxxxxxyyxxxx        (C) holds, (A) does not.
```

The remaining case, namely (A) and (D) fail to hold but (B) holds, is not possible because of the nature of m°. To see this, let $A_{i*} \neq \phi$ and $A_{i*+1} = \phi,\ldots,A_r = \phi$. If $i* = r$, then if B_{i*} is not separating, the interval consisting of the last element of A_{i*}, all the elements of B_{i*} and the first element of A_{r+1}, is a completely matchable interval. Thus we can construct a matching of larger cardinality than m°. If $i* < r$, then the interval consisting of the last element of A_{i*}, all the elements of B_{i*} and the first element of B_{i*+1}, is a completely matchable interval. Then, using Lemma 3.2.16, we may add the last element of A_{i*} to $X^+ + Y^+$ and take out the last element of B_{i*+1}, matching instead the elements of B_{i*+1} except for the first and last, to each other only. The result is a matching in M' which is lexicographically smaller than m°. This is a contradiction and so the remaining case cannot occur.

The value of k is crucial in determining what can happen to $m^\circ(X,Y,\Gamma_z)$ if $B_r(X,Y,\Gamma_z)$ is removed, in particular, whether the deficiency will change, as will be seen in the next section.

3.5 Recurrence equations for $K(u)$, $u \in C$.

Definition 3.5.1: Let $D_0(l,s)$ be the set of all non-switching intervals consisting of l elements of X and l elements of Y, which start with an element of X and end with s consecutive elements of Y. Let $D_1(l,x)$ be the set of members of $D_0(l,s)$ which are separating and let $D_2(l,s) = D_0(l,s)-D_1(l,s)$. Let $D_0'(l,s), D_1'(l,s)$, and $D_2'(l,s)$ be the corresponding sets of intervals which start with an element of Y. □

As an example, let $z = 2$ and $(X,Y) = \overset{----++++----}{\text{xxxxxxyyyyyy}}$. Then $\kappa(X,Y) = (0,6,4,2,4)$. In this case any element of $D_1'(3,1)$ may be inserted immediately after the 10th position to give an element (X',Y') of $\kappa^{-1}(1,9,4,1,2)$ such that $a(X',Y') = (X,Y)$. Distinct elements of $D_1'(3,1)$ yield distinct elements (X',Y'). Further, for any element (X',Y') in $\kappa^{-1}(1,9,4,1,2) \cap a^{-1}(\kappa^{-1}(0,6,4,2,4))$, it is clear that $B_r(X',Y',\Gamma_z) \in D_1'(3,1)$, because $B_r(X',Y',\Gamma_z)$ must be separating when $\kappa(X',Y') = 1$. It follows that there will be a 1-1 correspondence between $\kappa^{-1}(1,9,4,1,2) \cap a^{-1}(\kappa^{-1}(0,6,4,2,4))$ and $D_1'(3,1)$, i.e., $T(u,v) = \#D_1'(3,1)$, where $u = (0,6,4,2,4)$ and $v = (1,9,4,1,2)$. The general situation is the following.

<u>Proposition 3.5.2</u>: Let $(X,Y) \in \kappa^{-1}(u)$, with the last element in Y (for definiteness). Let $v \in C$ with $n(v) > n(u)$. Let $l = n(v) - n(u)$.

a) Suppose $d(u) = d(v) > 0$, and $t(u) > t(v)$. Then if $k(v) = 1$ (resp. $k(v) = 2$), $T(u,v) = \#D_1'(l,s(v))$ (resp. $\#D_2'(l,s(v))$). This corresponds to the previous example.

b) Let $d(u) = d(v) > 0$, and $t(u) = t(v) > 0$. Then if $k(u) = 0$ (resp. $k(u) = 1$) and $k(v) = 1$ (resp. $k(v) = 0$), $T(u,v) = \#D_0(l,s(v))$ (resp. $\#D_0'(l,s(v))$). If $t(u) = t(v) = 0$ and $d(u) = d(v) > 0$, then $T(u,v) = \#D_0(l,s,(v))$. For example, $(X,Y) = \overset{----++++----}{xxxxxxyyyyyy}$, insert an element of $D_0'(l,s(v))$ after the 8th position.

c) Let $d(u) = d(v) = 0$. Then if $k(u) = k(v) = 0$, $T(u,v) = \#D_0'(l,s(v))$. In this case (X,Y) is completely matched and we insert an element of $D_0'(l,s(v))$ (or $D_0(l,s(v))$ as appropriate) after the final position. The resulting (X',Y') will have $k(v) = 0$, since $d(v) = 0$.

d) Let $d(v) = d(u) + s(u)$ and $t(v) = t(u) + s(u)$ and $k(u) = k(v) = 0$. In this case $T(u,v) = \#D_1(l,s(v))$. For example if

$(X,Y) = \overset{----++++----}{xxxxxxyyyyyy}$, insert a separating interval after the 6th position, i.e., immediately after the last element of $B_r(X,Y) \cap X$.

e) If none of the above relationships holds between the parameters of u and v, then $a^{-1}(X,Y) \cap \kappa^{-1}(v) = \phi$, i.e., there is no element (X',Y') of X which is classified as v from which we can remove $B_r(X',Y',\Gamma_z)$ to get an element (X,Y) in $\kappa^{-1}(u)$. We postpone the proof of 3.5.2 until after the following corollary of 3.5.2 and 3.1.2.

<u>Proposition 3.5.3</u>:

$$K(0,n,d,s,t) = \sum_{=1}^{n-1} \#D_0(l,s) \sum_{i=1}^{z} K(1,n-l,d,i,t)$$

$$+ \sum_{l=1}^{n-1} \#D_1(l,s) \sum_{i=1}^{\min(z,t)} K(0,n-l,d-i,i,t-i)$$

$$+ (\text{if } d = 0)$$

$$\sum_{l=1}^{n} \#D_0(l,s) \sum_{i=0}^{z} K(0,n-l,0,i,0) .$$

(The first line includes classifications such that (b) of 3.5.2 holds; the second includes classifications such that (d) holds; and the third and fourth such that (c) holds.)

We will write below $K(\cdot,n,d,s,t) = K(0,n,d,s,t) + K(1,n,d,s,t) + K(2,n,d,s,t)$.

$$K(1,n,d,s,t) = \sum_{l=1}^{n-1} \#D_0(l,s) \sum_{i=0}^{z} K(0,n-l,d,i,t)$$

$$+ \sum_{l=1}^{n-1} \#D_1(l,s) \sum_{t'>t} \sum_{i=1}^{z} K(\cdot,n-l,d,i,t') .$$

(The first line comes from classifications such that (b) holds. The second comes from those for which (a) holds.)

$$K(2,n,d,s,t) = \sum_{l=1}^{n-1} \#D_2(l,s) \quad \sum_{t'>t} \sum_{i=1}^{z} K(\cdot,n-l,d,i,t') .$$

(Corresponds to the case when (a) holds.) □

Proof of 3.5.2.

(a) For u and v related by $k(v) = 1$, $d(u) = d(v)$, and $t(u) > t(v)$, we wish to show that for each $(X,Y) \in \kappa^{-1}(u)$, there is only one way to insert an interval into (X,Y) to get an element (X',Y') of $\kappa^{-1}(v)$ such that the inserted interval coincides with $B_r(X',Y',\Gamma_z)$, i.e., such that $a(X',Y') = (X,Y)$. This is to insert an element of $D_1'(l,s(v))$ into (X,Y) immediately in front of the last $t(v)$ elements of $A_{r+1}(X,Y,\Gamma_z)$. It will be apparent from the argument that any element of $D_1'(l,s(v))$ can be so inserted into (X,Y). Necessarily for distinct elements of $D_1'(l,s(v))$ inserted at the same location into (X,Y), the resulting elements (X',Y') are distinct. These statements hold independently of the choice of (X,Y) for any $(X,Y) \in \kappa^{-1}(u)$, given the choice of v. This will demonstrate a 1-1 correspondence beween $a^{-1}(X,Y) \cap \kappa^{-1}(v)$ and $D_1'(l,s(v))$.

Given the values $t(u)$ and $t(v)$, provided that the inserted element does actually correspond to $B_r(X,Y,\Gamma_z)$, it must have been inserted after the first $t(u) - t(v)$ elements of $A_{r+1}(X,Y,\Gamma_z)$ and it must be in $D_1'(l,s(v))$; if it is in $D_0(l,s(v))$, we would violate 3.2.14(b). If it is in $D_2'(l,s(v))$ we would violate (A) above, since by the next Lemma the inserted elements are in this case matched only to each other by $m^\circ(X,Y,\Gamma_z)$ (and hence $A_r(X',Y',\Gamma_z)$ is nonempty.)

Lemma 3.5.4: (i) Let (X,Y) partition $\{1,2,\ldots,n+n'\}$. Let $I = [j_0+1,\ j_0+2l]$ be a non-switching interval containing l elements of

X and l of Y, and starting with an element of Y (we could equally well choose X). If no elements of $X \cap [j_0+2l+1, n+n']$ are matched by m° to elements of $I \cap Y$, then no elements of $I \cap X$ are matched by m° to elements of $Y \cap [1, j_0]$. (For our purposes, we will apply the Lemma to $I = B_r(X', Y', \Gamma_z)$ -- the inserted interval.)

(ii) If no elements of $Y \cap [1, j_0]$ are matched by m° to elements of $I \cap X$, then no elements of $I \cap Y$ are matched by m° to elements of $Y \cap [j_0+2l+1, n+n']$.

<u>Proof</u>: Let x_j be the first element of $I \cap X$. Then $x_j > j_0+1$. If x_j is matched to j_0 or some earlier element of Y, then we must have $j_0+1 \in Y^+(m^\circ)$, since m° is compressed. But then j_0+1 must be matched to an element $x_{j'}$ of $I \cap X$ such that $x_{j'} > x_j$. Repeating this argument using $x_{j'}$ in place of x_j, it will follow that since I is non-switching, each element of $I \cap Y$ must be in $Y^+(m^\circ)$. But because m° matches X^+ to Y^+ in order, the elements of $I \cap Y$ can only be matched to elements of $I \cap X$ greater than x_j , since by hypothesis no elements of $X \cap [j_0+2l+1, n+n']$ are mapped to $I \cap Y$. This implies that $I \cap Y$ is matched to a proper subset of $I \cap X$, which is impossible. This contradiction establishes (i).

(ii) If the interval I is non-switching as defined above, as we run through the elements starting at the left-hand endpoint, then it is also "non-switching" as we run through the elements in reverse order, starting with the right-hand endpoint. (ii) is thus the same as (i), except that in the argument take the elements in reverse order, i.e, going in the direction $n+n', \ldots, 2, 1$. □

It follows that in case (a) the inserted elements can be matched only to each other and therefore will comprise $B_r(X',Y',\Gamma_z)$. Indeed, there are no elements of X after the inserted inverval, since it is inserted in the middle of $A_{r+1}(X,Y,\Gamma_z)$. This is true for intervals in $D_2'(l,s(v))$ as well as $D_1'(l,s(v))$, hence the result when $k(v) = 2$.

(b) We need a preliminary Lemma.

Lemma 3.5.5: Let (X,Y) partition $\{1,2,\ldots,n+n'\}$ into n elements of X and n' of Y. Suppose $n+n' \in Y^-(m°(X,Y,\Gamma_z))$. Let (X',Y') be formed by inserting j Y-type elements at the end of (X,Y). Then $m°(X,Y,\Gamma_z) = m°(X',Y',\Gamma_z)$.

Proof: It suffices to prove the result for $j = 1$; for $j > 1$ insert the elements one at a time, using the result for $j = 1$.

If $N(X',Y',\Gamma_z) = N(X,Y,\Gamma_z)$, then $m°(X,Y,\Gamma_z) = m°(X',Y',\Gamma_z)$ by Proposition 3.2.9. So assume $N(X',Y',\Gamma_z) = N(X,Y,\Gamma_z) + 1$. (It is easy to see that $N(X',Y',\Gamma_z) \le N(X,Y,\Gamma_z) + 1$, since only one Y-type element has been added to (X,Y).)

If $m°(X,Y,\Gamma_z) \neq m°(X',Y',\Gamma_z)$, then $n+n'+1 \in Y^+(m°(X',Y',\Gamma_z))$ and hence since $m° \in \mathcal{M}'$, $(x^*, n+n'+1) \in m°(X',Y',\Gamma_z)$, where x^* is the n-th and final element of X, and hence since $X = X'$, of X'. The matching m' defined as $m° - \{(x^*,n+n'+1)\}$ is then a maximal matching in $\mathcal{M}(X,Y,\Gamma_z)$ (but of course not in $\mathcal{M}(X',Y',\Gamma_z)$), and m' does not include x^* as an initial vertex. By Proposition 3.2.5, the minimal matching m in $\mathcal{M}(X,Y,\Gamma_z)$ also does not include x^* as an initial vertex. Then $[x^*+1,n+n'] \subseteq Y^+(m)$ since $|x^* - (n+n')| \le z$, and if one of those elements were in $Y^-(m)$, it could be matched to $x^* \in X^-(m)$ to increase the cardinality of m. In particular $n+n' \in Y^+(m)$.

By Proposition 3.2.11, $m^\circ(X,Y,\Gamma_Z) = m^C$ and therefore since m is lexicographically minimal in M and so there are no elements of $Y^-(m)$ in the interval $[\, m(n+n'), n+n']$ (where $m(n+n')$ denotes the initial vertex of the arc in m with terminal vertex $n+n'$), then in constructing m^C (see Definition 3.2.7), $n+n'$ will never be replaced by another element of Y and hence $n+n' \in Y^+(m^\circ)$. This contradicts our hypothesis. □

As in (a), if the inserted elements do indeed form $B_r(X',Y',\Gamma_Z)$ we can conclude that the insertion must be made at the location claimed, i.e., right after $B_r(X,Y,\Gamma_Z)$. It will thus suffice to show that the elements of an interval inserted at this location are mapped only to each other by $m^\circ(X',Y',\Gamma_Z)$.

Consider now the elements of (X,Y) up through and including $B_{i^*}(X,Y,\Gamma_Z)$. Suppose for definiteness that $A_{i^*} \subseteq X$, so that B_{i^*} begins with an element of X and ends with an element of Y. Then in (X',Y',Γ_Z), the element after $B_{i^*}(X,Y,\Gamma_Z)$ will be an element of Y' (either the first element of $B_{i^*+1}(X,Y,\Gamma_Z)$ or the first element of the inserted interval, depending on whether $i^* = r$). By 3.5.4 (ii), if this element of Y' were matched to an earlier element of X' by m°, some element of A_{i^*} would have to be also matched to an element of B_{i^*}, violating the assumption that B_{i^*} is separating. Hence by 3.5.4 (i) no Y'-type element of B_{i^*} or subsequent to B_{i^*} can be matched by m° to any element to the left of B_{i^*}. But by 3.5.5, since $A_{i^*} \neq \emptyset$, no element of $X \cap B_{i*}$ can be matched by m° to any elements to the left of B_{i*}. Hence the elements of $B_{i*}, B_{i*+1}, \ldots, B_r(X,Y,\Gamma_z)$ and the inserted interval are matched only to each other, this being the minimal compressed matching for the whole right-hand portion of the partition (X,Y) starting with B_{i*}.

The existence of a 1-1 correspondence between $a^{-1}(X,Y) \cap \kappa^{-1}(v)$ and either $D_0(l,s(v))$ (if $k(v) = 1$) or $D_0'(l,s(v))$ (if $k(v) = 0$) follows.

(c) In this case, both (X,Y) and (X',Y') are completely matched, so any completely matchable nonswitching interval starting with the right type of element can be inserted at the end of (X,Y).

(d) In the previous cases, the insertion of an interval has left the rest of the matching unchanged. In case (d) the situation is more complicated.

We first note that there are four possibilities for (X',Y'):

(i) $A_r(X',Y',\Gamma_z)$ and $A_{r+1}(X',Y',\Gamma_z)$ are of the same type, i.e., both $\subseteq X'$ or both $\subseteq Y'$, or $A_r \neq \emptyset$ and $A_{r+1} = \emptyset$.

(ii) $A_i = \emptyset$ for $i = 1,\ldots,r+1$.

(iii) $A_r = \emptyset$, but some earlier set $A_i \neq \emptyset$. A_{r+1} is arbitrary in this case.

(iv) A_r and A_{r+1} are of opposite type, and both are nonempty.

Case (ii) corresponds to (c) above. Case (i) corresponds to (a) and case (iii) corresponds to (b). Only in case (iv) have we not seen how the parameters of $\kappa(X',Y')$ and $\kappa(X,Y)$ are related. (Note: (e) will have been proved when we show that (iv) corresponds to (d)).

In case (iv), when $B_r(X',Y',\Gamma_z)$ is removed, some of the elements of $A_{r+1}(X',Y',\Gamma_z)$ will be matched to elements of $A_r(X',Y',\Gamma_z)$, and will be the last $s(u)$ elements of $B_r(X,Y,\Gamma_z)$. Hence the inserted interval which is assumed to become $B_r(X',Y',\Gamma_z)$ must be inserted in the location described in (d), i.e., just before the last $s(u)$ elements of $B_r(X,Y,\Gamma_z)$. The inserted interval must be separating if it is to coincide with $B_r(X',Y',\Gamma_z)$, because $A_{r+1}(X',Y',\Gamma_z)$ and $A_r(X',Y',\Gamma_z)$ are of opposite type. Further, we must have $k(u) = 0$ and $t(v) = t(u) + s(u)$. ((iv) implies that $k(v) = 0$, by 3.2.14 (d).)

It remains to be shown that $d(v) = d(u) + s(u)$ and that when a separating interval starting with the appropriate type element is inserted in the location described, the inserted elements are matched to each other. The following example shows what might happen when such an insert is made. Let $z = 3$.

$$(X,Y) = \overset{++++--++++++--}{xxyyxxxxyxyyyy}$$

$$(X',Y') = \overset{++++---++-++++++----}{xxyyxxxxyx\underline{xxxyyy}yyyy}$$

Suppose for definiteness that $B_r(X,Y,\Gamma_z)$ starts with an element of X, so that $A_{r+1}(X,Y,\Gamma_z) \subseteq Y$. Let I denote the interval of elements of $B_r(X,Y,\Gamma_z)$ up to and including the last element of $B_r \cap X$. We argue

(1) Since the inserted interval, which we will call J, is separating, no element of $Y \cap J$ can be matched by $m^\circ(X',Y',\Gamma_z)$ to an element earlier than J. This follows from 3.5.4 (ii), as shown in (b) above, since $X_n A_{r+1}(X',Y',\Gamma_z) = \emptyset$.

(2) Consequently, as in 3.5.4(i), we can see that no element of $I \cap Y$ can be matched to an element earlier than I.

(3) If $A_r(X,Y,\Gamma_z) \neq \emptyset$, then it follows by 3.5.5 that elements of

$B_r(X,Y,\Gamma_z)$ and later are matched only to each other. Thus the elements of J, the $\#B_r(X,Y,\Gamma_z)/2 - s(u)$ elements of $Y \cap I$ and the $\#B_r(X,Y,\Gamma_z)/2$ elements of $X \cap I$ are mapped only to each other by $m°(X',Y',\Gamma_z)$, and hence the deficiency of $m°(X',Y',\Gamma_z)$ is increased by $s(u)$ over that of $m°(X,Y,\Gamma_z)$ (the part of the matching earlier than B_r remains unchanged.) Since $m°$ is the minimal compressed matching, it is clear that the last element of $X \cap I \in X^-(m°(X',Y',\Gamma_z))$, and that the elements of the inserted interval are matched only to each other.

(4) If $A_r(X,Y,\Gamma_z) = \emptyset$, then the sets $B_{i*},B_{i*+1},\ldots,B_r$ are adjacent to each other. But in this case as in (b) above, since B_{i*} is separating, no elements of B_{i*} or later elements can be matched by $m°$ to elements of A_{i*} or earlier, and hence the elements of $B_{i*},B_{i*+1},\ldots B_{r-1}$ are matched only to each other, giving the same conclusion as in (3).

□

3.6 Reducing the recurrence equations.

The following result places limits on the range of the indices of summation in the equations of Proposition 3.5.2.

Proposition 3.6.1: Let (X,Y) partition $\{1,2,\ldots,2n\}$, where $n \geq 2$. Then $D(X,Y,\Gamma_z) \leq n - z$, i.e., $N(X,Y,\Gamma_z) \geq z$, $1 \leq z \leq n$.

Proof: Suppose $N(X,Y,\Gamma_z) < z$. Then $X^+(m°) + Y^+(m°)$ must contain at least one separating interval in order to keep the elements of X^- and Y^- from being matched to each other. Using 3.2.16, we can, as before, show that there must be a non-switching separating interval, I. Since $N(X,Y,\Gamma_z) < z$, $\#(I) < 2z$.

However, a non-switching interval I of l elements of X and l elements of Y cannot be separating if $l < z$. To see this, let the elements of $I \cap X$ be $a_1 < \ldots < a_l$ and those of $I \cap Y$ be $b_1 < \ldots < b_l$, for $l < z$. Suppose for definiteness that $a_1 < b_1$. For each i, $2 \le i \le l$, there are at least $i - 2$ elements of I to the left of a_i and at least $l - i$ to the right of b_i. So $|a_{i-1} - b_i| \le 2l - (l-i) - (i-2) - 1 \le l + 1 \le z$. Hence $(a_{i-1}, b_i) \in \Gamma_z$. If we insert an X-type element in front of I and a Y-type element at the end of I, then we can match (in Γ_z) b_1 to the new X-type element and a_l to the new Y-type element. Hence I is not separating. □

We can reduce the number of variables appearing in 3.5.2 by collecting terms in the following manner.

Define
$$V_1(n,d,t) = \sum_{t'>t} \sum_{i=1}^{z} K(\cdot,n,d,i,t')$$

$$V_2(n,d,t) = \sum_{i=1}^{z} K(0,n,d,i,t)$$

$$V_3(n,d,t) = \sum_{i=1}^{z} K(1,n,d,i,t)$$

$$V_4(n,d,t) = \sum_{i=1}^{\min(z,t)} K(0,n,d-i,i,t-i) \quad .$$

The equations of 3.5.2 then become:

$$\begin{aligned} K(0,n,d,s,t) = & \sum_{l=1}^{n-1} \#D_0(l,s)V_3(n-l,d,t) \\ & + \sum_{l=1}^{n-1} \#D_1(l,s)V_4(n-l,d,t) \\ & + \textit{(if } d=0\textit{)} \sum_{l=0}^{n} \#D_0(l,s)V_2(n-l,0,0) \quad . \end{aligned}$$

$$K(1,n,d,s,t) = \sum_{l=1}^{n-1} \#D_0(l,s)V_2(n-l,d,t)$$

$$+ \sum_{l=1}^{n-1} \#D_1(l,s)V_1(n-l,d,t) \ .$$

$$K(2,n,d,s,t) = \sum_{l=1}^{n-1} \#D_2(l,s)V_1(n-l,d,t) \ .$$

Our initial conditions are $V_2(0,0,0) = 1$, corresponding to the case $n = 0$ and $(X,Y) = \emptyset$, and $V_1(0,0,0) = 0$, $V_3(0,0,0) = 0$, $V_4(0,0,0) = 0$. If $V_i(p,d,t)$ are known for $i=1,2,3,4$ and all d and t and all $p \le n-1$, for some n, then once $K(k,n,d,s,t)$ have been computed for all k,d,s, and t, we can compute $V_i(n,d,t)$ for $i=1,2,3,4$ and all values of d and t. Thus in evaluating the recurrence equations, it will never be necessary to store the values of $K(n,k,d,s,t)$. This reduction from four indices to three makes evaluation by electronic computer feasible for values of n up to several hundred, once the values of $\#D_i(l,s)$ are known.

3.7 Computation of $\#D_i(l,s)$.

Using an argument based on 3.1.2 we will be able to find $\#D_i(l,s)$, $i=0,1,2$, for values of l up to several hundred and for values of z up to 12. The limit of $z = 12$ is established by computer storage facilities. The method will suffice to construct tables of our distribution function up to about $n = 70$, after which we cannot compute the upper tail.

After Sections 3.1 - 3.6, we have remaining only the problem of enumerating for a given value of z, the completely matchable non-switching partitions (X,Y) of $\{1,2,\ldots,2l\}$ such that $\#(X) = \#(Y)$ and such that the interval is separating, and also those such that the interval is non-separating. The "separating" question will be difficult to handle directly. However, we can instead enumerate $\#D_0(l,s)$ and $\#D_3(l,s)$, where $D_3(l,s)$ is defined as the subset of $D_0(l,s)$ consisting of strictly non-switching partitions.

Proposition 3.7.1: $\#D_2(l,s) = \#D_3(l+1,s+1)$.

Proof: For $(X,Y) \in D_2(l,s)$, let $t(X,Y)$ be the partition of $\{1,2,\ldots,2(l+1)\}$ formed by inserting an X-type element at the beginning of (X,Y) and a Y-type element at the end. Then, since (X,Y) is non-separating, $t(X,Y) \in D_3(l+1,s+1)$. t is clearly an injective mapping onto a subset of $D_3(l+1,s+1)$. In fact, t is onto all of $D_3(l+1,s+1)$, with t^{-1} being defined by removing the first and last elements of $(X',Y') \in D_3(l+1,s+1)$. So defined, $t^{-1}(X',Y')$ is completely matchable, by 3.2.16, and is clearly non-switching (though not necessarily strictly non-switching). $t^{-1}(X',Y')$ is not separating since (X',Y') is completely matchable. Hence $t^{-1}(X',Y') \in D_2(l,s)$. □

If we can find $\#D_0(l,s)$ and $\#D_3(l,s)$ for all l and s, then we find $\#D_2(l,s) = \#D_3(l+1,s+1)$ and $\#D_1(l,s) = \#D_0(l,s) - \#D_2(l,s)$.

Let us first address the problem of finding for a given value of z, $\#D_0(l,s)$ for all l and s. We use the ideas of 3.1.2. Let $\mathcal{Y}_l$ be the class of all non-switching completely matchable partitions of $\{1,2,\ldots,2l\}$ such that $\#(X) = \#(Y)$ and $1 \in X$. Let $\mathcal{Y} = \bigcup_{l=0}^{\infty} \mathcal{Y}_l$.

Define the mapping $a\colon \mathcal{Y}_l \to \mathcal{Y}_{l-1}$ (for $l \geq 1$) by letting $a(X',Y')$ be the element of $\mathcal{Y}_{l-1}$ obtained by removing from (X',Y') the last element of X' and the last element of Y'. We let $\mathcal{Y}_0 = \{\emptyset\}$ and $a(\emptyset) = \emptyset$.

Define the classification set C for elements of $\mathcal{Y}$ as follows. Let $(X,Y) \in \mathcal{Y}_l$. Suppose (X,Y) ends with s successive elements of $\mathcal{Y}$, namely $y_{l-s+1}, y_{l-s+2}, \ldots, y_l$. Define $r_1,\ldots,r_s$ by $r_1 = |x_{l-s+1} - y_{l-s+1}|$, $r_2 = |x_{l-s+2} - y_{l-s+2}|,\ldots,r_s = |x-y_l|$. For example, if $z = 4$ and $(X,Y) = xxyxxyxyyy$, then $l = 5$, $s = 3$, $r_1 = 4$, $r_2 = 4$, $r_3 = 3$.) Then let $(X,Y) = (l,s,r_1,\ldots,r_s)$.

Proposition 3.7.2: Let $u = (l,s,r_1,\ldots,r_s)$ be a vector of positive integers. The following conditions are necessary and sufficient for $u \in C$.

(1) $l \geq s$, $s \leq z$.

(2) $r_i \leq z$, for $1 \leq i \leq s-1$, and $r_s = s$.

(3) $l \geq r_1 \geq r_2 \geq \ldots \geq r_s$.

Proof: The necessity of (1) and (2) are obvious as is the necessity of $r_1 \geq r_2 \geq \ldots \geq r_s$. $l \geq r_1$, since between x_{l-s+1} and y_{l-s+1} there are $s-1$ elements of X and $r_1 - r_s$ elements of Y. So we have located $r_1 - r_s$ elements of Y not including $y_{l-s+1},\ldots,y$. Hence $r_1 \leq \#(Y) = l$.

To show sufficiency, we describe an algorithm by which an element $(X,Y) \in \mathcal{Y}_l$ can be constructed with u as its classification vector. Start with an X-type element, followed by s Y-type elements. Insert in front of these elements an X-type element followed by $r_{s-1} - r_s$ Y-type elements. In front of these elements, insert an X-type element followed by $r_{s-2} - r_{s-1}$ Y-type elements,..., in front of these elements insert an X-type element followed by $r_1 - r_2$ Y-type elements. We now have assembled r_1 Y-type elements and s X-type elements. Insert $r_1 - s$ X-type elements in front of these. If $l > r_1$, insert an element of $\mathcal{Y}_{l-r_1}$ at the beginning. The resulting partition of $\{1,2,\ldots,2l\}$ is clearly non-switching and has classification vector u. We need only show that it is completely matchable. This follows from (2): clearly the initial element of $\mathcal{Y}_{l-r_1}$ is completely matchable (to itself); the first s X-type elements inserted, one at a time, by the algorithm can be matched to the right-most s Y-type elements; since x_{l-s+1} and y_{l-s+1} are matched, then the r_1-s elements of X inserted at the same time (which will be immediately in front of x_{l-s+1}) can be matched respectively to $y_{l-r_1+1},\ldots,y_{l-s}$. □

For $(X',Y') \in \mathcal{Y}$, $u,v \in C$, we have $\#a(X',Y') = 1$. For (X,Y) in $\kappa^{-1}(v)$, we may obtain every element of $a^{-1}(X,Y)$ by inserting an X-type element after one of the last s(u) elements of Y and inserting a Y-type element at the end. It will be clear that for each $v \in C$, $T(u,v) = \#a^{-1}(X,Y) \cap \kappa^{-1}(v) = 0$ or 1, depending on u.

Proposition 3.7.3: Let $u,v \in C$. The following are necessary and sufficient conditions for $T(u,v) = 1$.

(1) $l(v) = l(u) + 1$

(2) $s(v) \leq s(u) + 1$

(3) For $i=1,2,\ldots,s(v)-1$, we have $r_{s(v)-i+1}(u) + 1 = r_{s(v)-i}(v)$.

Proof: (1) and (2) are clearly necessary (consider the definition of a). Note that in performing the insertion described to get an element $(X',Y') \in a^{-1}(X,Y) \cap \kappa^{-1}(v)$, for each of the elements $y_l, y_{l-1}, \ldots, y_{l-s(v)+1}$ of (X,Y), we have inserted exactly one X-type element between that element and the corresponding element $x\ , x_{l-1}, \ldots,$ or $x_{l-s(v)+1}$ of X. Then given $(X,Y) \in \kappa^{-1}(u)$, we can make the appropriate insertion to obtain an element $(X',Y') \in a^{-1}(X,Y) \cap \kappa^{-1}(v)$, if and only if (1),(2) and (3) are satisfied. □

Proposition 3.7.4: Let $l \geq z$. Let $U = \{u \in C: s(u) = s_0\}$ for some s_0, $1 \leq s_0 \leq \min(l,z)$. Then $\#(U) = \sum_{j=1}^{\min(l,z)-s_0} \binom{j+s_0-2}{j}$.

Proof: $\#(U) = \sum_{j=s_0}^{z} \#\{u \in C: s(u) = s_0 \text{ and } r_1(u) = j\}$.

Thus for given j, $s_0 \leq j \leq z$, we wish to evaluate $\#(U_j)$, where $U_j = \{u \in C: s(u) = s_0,\ r_1(u) = j\}$. Let $u = (l, s_0, j, r_2, \ldots, r_{s_0})$. Let $j_1 = j - r_2$, $j_2 = r_2 - r_3, \ldots, j_{s_0-1} = r_{s_0-1} - s_0$. Then

$$(*) \qquad \sum_{i=1}^{s_0-1} j_i = j - s_0\ ; \quad \text{and} \quad j_1 \geq 0, \ldots, j_{s_0-1} \geq 0 .$$

It is clear that, conversely, for any integers $j_1, \ldots, j_{s_0-1}$ which satisfy (*), the vector

$$(l, s_0, j, s_0 + \sum_{i=2}^{s_0-1} j_i, s_0 + \sum_{i=3}^{s_0-1} j_i, \ldots, s_0 + j_{s_0-1}, s_0) \in U_j ,$$

by Proposition 3.7.2. It follows that $\#(U_j)$ is the number of sequences $j_1, \ldots, j_{s_0-1}$ satisfying (*). This number is $\binom{j+s_0-2}{j}$, (see, for example, [13]).

If $l < z$, then we need only consider j such that $s_0 \le j \le l$.

Using the notation of 3.1.2, □

$$\#D_0(l_0,s_0) = \sum_{\{u \in C: s(u)=s_0, l(u)=l_0\}} K(u).$$

Once the values of $T(u,v)$ are known for all u and v, we can find $K(u)$ for all $u \in C$ by starting with $K(\kappa^{-1}(\emptyset)) = 1$. We will not actually compute the values $T(u,v)$, but will use them implicitly as follows.

Let C_s be the set of all possible values of $(r_1(u),\ldots,r_s(u))$ where $u \in C$ such that $s(u) = s$. For two elements $w = (r_1,\ldots,r_s)$ and $w' = (r_1',\ldots,r_s')$ of C_s, we say that $w < w'$ when the vector $(r_1,\ldots,r_s)$ is lexicographically smaller than $(r_1',\ldots,r_s')$. For each $u \in C$, let $j(u)$ denote the rank of $(r_1,\ldots,r_{s(u)})$ in $C_{s(u)}$ in this lexicographic ordering. A classification $u \in C$ can now be described by three parameters, $(l(u),s(u),j(u))$. Let $v \in C$ and let $j^*(j(v),s_0 s(v))$ be the rank in C_{s_0} of that vector $(r_1,\ldots,r_{s_0})$ such that $u = (l(v)-1,s_0,r_1,\ldots,r_{s_0})$. By 3.7.3 it can be seen that there is at most one such vector; if $T(u,v) = 0$, there will be no such vector. Then

$$K(l(v),s(v),j(v)) = \sum_{s=s(v)-1}^{z} K(l(v)-1,s,j^*(j(v),s,s(v))),$$

with the convention that $K(0,0,0) = 1$ and $K(l,s,0) = 0$ otherwise.

The array $j^*(j',s,s')$ for $1 \le s' \le z$ and $s'-1 \le s \le z$ and $1 \le j' \le \#(U_{s'})$ ($U_{s'}$ as in 3.7.4) was determined using 3.7.3 essentially by brute force, though systematically. The pattern established by the lexicographic ordering permits this. The practical limitation on the value of z which can be handled is due to the size of this array.

To evaluate $\#D_3(l,s)$, we use the same procedure except that in 3.7.2 we have the additional condition that $s > 1$ if $l > 1$. To see this, consider the sequence $a(X,Y)$, $a^2(X,Y),\ldots,a^{l-1}(X,Y),\emptyset$. If (X,Y) is strictly non-switching, then no element of this sequence except $a^{l-1}(X,Y)$ may be a partition which ends with ....xy, i.e., such that $s(u) = 1$. It can then be seen that $s > 1$ if $l > 1$ is also a sufficient condition for $(X,Y) \in D_3$.

The problem of enumerating $D_0(l,s)$ can be rephrased as a problem of enumerating the set of matrices of 0's and 1's which satisfy certain conditions. However, this equivalent problem does not produce a more satisfactory solution, and we will simply give the statement of this problem.

To enumerate the set of $l \times l$ matrices M such that:

(1) M consists only of 0's and 1's;

(2) The non-zero elements in each row are consecutive, with the last 1 being on the main diagonal;

(3) For $i=2,\ldots,l$, the first non-zero element in row i never appears in an earlier column than the first element of row i-1;

(4) The sum of the elements in each row is less than or equal to z.

Equivalently we may replace (4) by:

(5) M is symmetric;

(6) For $i=1,\ldots,l$, the total number of non-zero elements in the i-th row and i-th column (no element being counted twice) is less than or equal to z.

CHAPTER 4

Asymptotic Distribution of the Test Statistics

4.1 Asymptotic distribution for $d_{\mathcal{P}}$.

We use results from Sections 1.3 and 1.4 to get some preliminary lemmas. Let $\mathcal{F}$ be defined as in 1.3.9. If $F \in \mathcal{F}$ and $F = [b_1,c_1]\cup\ldots\cup[b_k,c_k]$, $0 \le b_1 \le c_1 < b_2 \le c_2 < \ldots < b_k \le c_k \le 1$, then the intervals $[b_i,c_i]$ will be referred to as the "component" intervals of F. For each $F \in \mathcal{F}$, we are interested in the total number of left-hand endpoints, excluding 1, of component intervals, plus the number of right-hand endpoints, excluding 0. Specifically we define

$$q(F) = \lim_{\delta \downarrow 0} \frac{\mu(F^{\delta}-F)}{\delta} ,$$

where μ denotes uniform measure on $[0,1]$. Let $\mathcal{F}^{(2\delta)}$ be defined as in 1.3.9. Note that $\mathcal{F}^{(2\delta)} = \{F \in \mathcal{F}: q(F) = q(F^{\delta]})\}$.

Define the function f on $D[0,1]$ by

$$f(y) = \sup_{F\in\mathcal{F}} \frac{1}{q(F)+1} \int_F dy(\mathbf{t}) , \quad \text{for all } y \in D[0,1].$$

(Note: for $F = \bigcup_{i=1}^{k} [b_i,c_i]$, we define $\int_F dy(t) = \sum_{i=1}^{k} (y(c_i) - y(b_i)).$)

For each $\delta > 0$, define the function $f^{(\delta)}$ on $D[0,1]$ by

$$f^{(\delta)}(y) = \sup_{F \in \mathcal{F}^{(2\delta)}} \frac{1}{q(F)+1} \int_F dy(t) .$$

Lemma 4.1.1: Let X_n and X be probability measures on $D[0,1]$ with $X(C[0,1]) = 1$, and let $X_n \Rightarrow X$. Let $\{\alpha_n\}$ be a sequence of numbers such that $\alpha_n \downarrow 0$. Then $f^{(\alpha_n)}(X_n) \Rightarrow f(X)$.

Proof: By Proposition 1.2.4, since $X(C[0,1]) = 1$, it suffices to show that for $y \in C[0,1]$, $y_n \to y$ in $D[0,1]$ implies that $f^{(\alpha_n)}(y_n) \to f(y)$.

We first show that $f(y) \leq \limsup_{n\to\infty} f^{(\alpha_n)}(y_n)$, i.e., that given $\varepsilon > 0$, for n sufficiently large,

$$\sup_{F\in\mathcal{F}} \frac{1}{q(F)+1} \int_F dy(t) \leq \sup_{F\in\mathcal{F}^{(\alpha_n)}} \frac{1}{q(F)+1} \int_F dy_n(t) + \varepsilon .$$

This will be done by showing that for n sufficiently large depending only on ε and on y, then for each $F \in \mathcal{F}$, there exists a set $G \in \mathcal{F}^{(\alpha_n)}$ such that

$$\frac{1}{q(F)+1} \int_F dy(t) \leq \frac{1}{q(G)+1} \int_G dy_n(t) + \varepsilon .$$

G will be constructed from F as in Proposition 1.3.10. I.e., if $F = \bigcup_{i=1}^{k} [b_i, c_i]$ as above, then G is defined as the union of F with every interval $[c_i, b_{i+1}]$ such that $b_{i+1} - c_i \leq 2\alpha_n$ and with $[0, b_1]$ if $b_1 \leq \alpha_n$ and with $[c_m, 1]$ if $c_m \geq 1 - \alpha_n$. Then clearly $G \in \mathcal{F}^{(2\alpha_n)}$.

Remark 1: $q(F) \geq q(G)$. (Clearly $G \in \mathcal{F}$ so $q(G)$ is defined.) Indeed, each component interval of G contains a point of some component interval of F and hence must contain the whole component interval. Each component interval of F intersects precisely one component interval of

G. If a component interval of G contains a component interval of F which has an endpoint at 0 (resp. 1), then that component interval of G has an endpoint at 0 (resp. 1). Thus the endpoints in (0,1) of component intervals of G can be placed in 1-1 correspondence with the endpoints in (0,1) of a subset of the component intervals of F. (Of course, a component interval of G may contain several component intervals of F.)

Remark 2: G - F consists of at most q(F) intervals each of length $< 2\alpha_n$. (These will be referred to as component intervals although in general they are not closed intervals and $G - F \not\subset F$.) Indeed, it is evident that for each component interval of G - F, at least one of its endpoints must also be the endpoint of a component interval of F. If that endpoint is 0 (resp. 1), then we must have $b_1 = c_1 = 0$ (resp. $b_1 = c_1 = 1$), in which case the endpoint is counted once, as a right-hand (resp. left-hand) endpoint. Therefore the number of component intervals of G - F is less than or equal to q(F).

Choose n sufficiently large that

$$\sup_{\substack{t_1,t_2\in[0,1] \\ |t_1-t_2|<2\alpha_n}} |y(t_1) - y(t_2)| < \frac{\varepsilon}{3} \quad \text{and} \quad \sup_{0\le t\le 1} |y_n(t) - y(t)| < \frac{\varepsilon}{3} .$$

This is possible because $y \in C[0,1]$ implies that y is uniformly continuous and further that convergence to y in D[0,1] is equivalent to convergence in the uniform metric.

But $$\left|\frac{1}{q(F)+1}\int_F dy(t) - \frac{1}{q(F)+1}\int_G dy_n(t)\right| \le$$

$$\frac{1}{q(F)+1}\left|\int_G dy_n(t) - \int_G dy(t)\right| + \frac{1}{q(F)+1}\left|\int_G dy(t) - \int_F dy(t)\right| .$$

By remark 2,

$$\left|\int_G dy - \int_F dy\right| \le 2q(F) \sup_{\substack{t_1,t_2\in[0,1] \\ |t_1-t_2|<2\alpha n}} |y(t_1) - y(t_2)| \le 2q(F)\frac{\varepsilon}{3} .$$

$$\left|\int_G dy_n - \int_G dy\right| \le q(G) \sup_{0\le t\le 1} |y_n(t) - y(t)| < q(G)\frac{\varepsilon}{3} \le q(F)\frac{\varepsilon}{3} .$$

Therefore,

$$\left|\frac{1}{q(F)+1}\int_F dy(t) - \frac{1}{q(F)+1}\int_G dy_n(t)\right| \le \varepsilon .$$

Hence,

$$(*) \qquad \frac{1}{q(F)+1}\int_F dy(t) \le \frac{1}{q(F)+1}\int_G dy_n(t) + \varepsilon .$$

We have two cases: (i) $\int_G y_n(t) \ge 0$ and (ii) $\int_G y_n(t) < 0$. In case (ii), replace G by $G' = \emptyset$. Then $G' \in F^{(2\alpha_n)}$. (ii) and (*) imply that

$$\frac{1}{q(F)+1}\int_F dy(t) \le \varepsilon = \frac{1}{q(G')+1}\int_{G'} dy_n(t) + \varepsilon.$$

In case (i), (*) and Remark 2 imply that

$$\frac{1}{q(F)+1}\int_F dy(t) \le \frac{1}{q(G)+1}\int_G dy_n(t) + \varepsilon$$

In either case we thus have the necessary inequality, showing that $f(y) \le \limsup_{n\to\infty} f^{(\alpha_n)}(y_n)$.

We complete the proof by showing that $\limsup_{n\to\infty} f^{(\alpha_n)}(y_n) \le f(y)$.

As above, given ε and choosing n sufficiently large, then for any $F \in \mathcal{F}$,

$$\left| \int_F dy_n(t) - \int_F dy(t) \right| \le q(F)\varepsilon ,$$

so

$$\int_F dy_n(t) \le \int_F dy(t) + q(F)\varepsilon .$$

Consequently,

$$\sup_{F\in\mathcal{F}^{(2\alpha_n)}} \frac{1}{q(F)+1} \int_F dy_n(t) \le \sup_{F\in\mathcal{F}^{(2\alpha_n)}} \frac{1}{q(F)+1} \int_F dy(t) + \varepsilon$$

$$\le \sup_{F\in\mathcal{F}} \frac{1}{q(F)+1} \int_F dy(t) + \varepsilon = f(y) + \varepsilon .$$

Since ε is arbitrary, $\limsup_{n\to\infty} f^{(\alpha_n)}(y_n) \le f(y)$. □

Recall from Proposition 1.3.9 that

$$\Pr\{d_{\mathcal{P}}(P,Q) \le c\} = \Pr\{ \sup_{F\in\mathcal{F}^{(2c)}} P(F) - Q(F^c) \le c\}$$

$$= \Pr\{P(F) \le Q(F^{c]}) + c; \quad \text{for all } F\in\mathcal{F}^{(2c)}\}$$

Letting μ denote uniform measure on $[0,1]$, let μ_m and ν_n be the empirical measures of the sets $\{r_i/N\}$ and $\{s_i/N\}$, where $r_i,\ldots,r_m$ and $s_1,\ldots,s_n$ are the ranks among the pooled observations of two i.i.d. random samples from μ, and $N = m+n$. Let γ_N be defined as $\frac{m}{N}\mu_m + \frac{n}{N}\nu_n$. In other words, γ_N is the measure which places mass $\frac{1}{N}$ at each of the points $1/N, 2/N,\ldots,1$.

<u>Proposition 4.1.2</u>: $n^{\frac{1}{2}}d_{\mathcal{P}}(\mu_n,\mu) \Rightarrow f(W^\circ)$ as $n\to\infty$.

Proof: Let c be a continuity point of the distribution of $f(W°)$.

$$\Pr\{d_P(\mu_n,\mu) \le cn^{-\frac{1}{2}}\} = \Pr\{\mu_n(F) \le \mu(F^{cn^{-\frac{1}{2}}}]) + cn^{-\frac{1}{2}} \text{for all } F \in F^{(2cn^{-\frac{1}{2}})}\}$$

$$= \Pr\{n^{\frac{1}{2}}(\mu_n(F) - \mu(F)) \le n^{\frac{1}{2}}\mu(F^{cn^{-\frac{1}{2}}}] - F) + c; \text{ for all } F \in F^{(2cn^{-\frac{1}{2}})}\} .$$

For $F \in F^{(2cn^{-\frac{1}{2}})}$, $\mu(F^{cn^{-\frac{1}{2}}}] - F) = q(F)cn^{-\frac{1}{2}}$, since for such a set F there will be no overlap at the ends of the component interval. Hence the last expression is equal to

$$\Pr\{n^{\frac{1}{2}}(\mu_n(F) - \mu(F)) \le (q(F)+1)c; \text{ for all } F \in F^{(2cn^{-\frac{1}{2}})}\}$$

$$= \Pr\Big\{ \sup_{F \in F^{(2cn^{-\frac{1}{2}})}} \frac{1}{q(F)+1} \int_F dX_n(t) \le c\Big\} = \Pr\{f^{(cn^{-\frac{1}{2}})}(X_n) \le c\} ,$$

where $X_n(t) = n^{\frac{1}{2}}(\mu_n[0,t] - \mu[0,t])$. By 1.4.2, $X_n \Rightarrow W°$ and hence by 4.1.1 $f^{(cn^{-\frac{1}{2}})}(X_n) \Rightarrow f(W°)$.

Hence

$$\lim_{n\to\infty} \Pr\{n^{\frac{1}{2}}d_P(\mu_n,\mu) \le c\}$$

$$= \lim_{n\to\infty} \Pr\{f^{(cn^{-\frac{1}{2}})}(X_n) \le c\}$$

$$= \Pr\{f(W°) \le c\} . \qquad \square$$

Proposition 4.1.3: $\left(\frac{n}{mN}\right)^{-\frac{1}{2}} d_P(\mu_m,\gamma_N) \Rightarrow f(W°)$ as $m,n \to \infty$

Proof: We treat separately the cases (i) $0 < a \le \frac{m}{n} \le b < \infty$, for some constants a and b; (ii) $\frac{m}{n} \to 0$; (iii) $\frac{m}{n} \to \infty$. This will suffice to show that $\left(\frac{n}{mN}\right)^{-\frac{1}{2}} d_P(\mu_m,\gamma_N) \Rightarrow f(W°)$ as $m,n \to \infty$; indeed, any subsequence (m_i,n_i) must contain a sub-subsequence satisfying (i), (ii), or (iii), so weak convergence would hold for that sub-subsequence.

(i) Let $\delta(m,n) = \left(\frac{n}{mN}\right)^{\frac{1}{2}}$. Let c be a continuity point of the distribution of $f(W°)$. By 1.3.9,

$$\Pr\{d_{\mathcal{P}}(\mu_m,\gamma_N) \le c\delta(m,n)\} = \Pr\{\mu_m(F) \le \gamma_N(F^{c\delta(m,n)]})$$

$$+ c\delta(m,n); \text{ for all } F\in\mathcal{F}^{(2c\delta(m,n))}\}$$

$$= \Pr\{\mu_m(F) - \gamma_N(F) \le \gamma_N(F^{c\delta(m,n)]} - F) + c\delta(m,n); \text{ for all } F\in\mathcal{F}^{(2c\delta(m,n))}\}$$

$$= \Pr\{\tfrac{n}{N}(\mu_m(F) - \nu_n(F)) \le q(F)(c\delta(m,n) + \eta(F))$$

$$+ c\delta(m,n); \text{ for all } F\in\mathcal{F}^{(2c\delta(m,n))}\},$$

where $\eta(F)$ satisfies $-\frac{1}{N} \le \eta(F) \le \frac{1}{N}$, since for any interval $I \subseteq [0,1]$, $|\gamma_N(I) - \mu(I)| \le \frac{1}{N}$. Thus $\eta(F) = 0\left(\frac{1}{N}\right)$, uniformly in F.

Multiplying through by $\delta(m,n)^{-1}$, we get

$$\Pr\{\left(\tfrac{mn}{N}\right)^{\frac{1}{2}}(\mu_m(F) - \nu_n(F)) \le (q(F)+1)c$$

$$+ q(F)\left(\tfrac{m}{n}\right)^{\frac{1}{2}} 0(N^{-\frac{1}{2}}); \text{ for all } F\in\mathcal{F}^{(2c\delta(m,n))}\}$$

$$= \Pr\{\sup_{F\in\mathcal{F}^{(2c\delta(m,n))}} \frac{1}{q(F)+1}\int_F dY_{m,n}(t) \le c + \left(\tfrac{m}{n}\right)^{\frac{1}{2}} 0(N^{-\frac{1}{2}});$$

$$\text{for all } F\in\mathcal{F}^{(2c\delta(m,n))}\}$$

$$= \Pr\{f^{(c\delta(m,n))}(Y_{m,n}) \le c + \left(\tfrac{m}{n}\right)^{\frac{1}{2}} 0(N^{-\frac{1}{2}})\},$$

where $Y_{m,n}$ is defined in 1.4.2. By 1.4.2, as $m,n \to \infty$ in such a way that (i) is satisfied, $Y_{m,n} \Rightarrow W°$, so by 4.1.1, $f^{(c\delta(m,n))}(Y_{m,n}) \Rightarrow f(W°)$. Since we have assumed $0 < a \le \frac{m}{n}$, $\left(\frac{m}{n}\right)^{\frac{1}{2}} 0(N^{-\frac{1}{2}}) = 0(N^{-\frac{1}{2}})$. Hence since c is a continuity point of the distribution of $f(W°)$,

$$\lim_{m,n\to\infty} \Pr\{\delta(m,n)^{-1} d_{\mathcal{P}}(\mu_m,\gamma_N) \le c\} = \Pr\{f(W^\circ) \le c\} .$$

(ii) In this case we consider

$$\Pr\{d_{\mathcal{P}}(\mu_m,\gamma_N) \le cm^{-\frac{1}{2}}\} = \Pr\{\mu_m(F) - \mu(F) \le (\gamma_N(F^{cm^{-\frac{1}{2}}}]) - \mu(F)) + cm^{-\frac{1}{2}}; \text{ for all } F\in\mathcal{F}^{(2cm^{-\frac{1}{2}})}\}$$

$$= \Pr\{\mu_m(F) - \mu(F) \le (\gamma_N(F^{cm^{-\frac{1}{2}}}]) - \mu(F^{cm^{-\frac{1}{2}}}])) + \mu(F^{cm^{-\frac{1}{2}}}] - F) + cm^{-\frac{1}{2}}; \quad \forall F\in\mathcal{F}^{(2cm^{-\frac{1}{2}})}\}$$

$$= \Pr\{m^{\frac{1}{2}}(\mu_m(F) - \mu(F)) \in q(F)(c+m^{\frac{1}{2}}\eta(F)) + c; \text{ for all } F\in\mathcal{F}^{(2cm^{-\frac{1}{2}})}\} ,$$

where $-\frac{1}{N} \le \eta(F) \le \frac{1}{N}$,

$$= \Pr\{ \sup_{F\in\mathcal{F}^{(2cm^{-\frac{1}{2}})}} \frac{1}{q(F)+1} \int_F d\mathcal{X}_m(t) \le c + m^{\frac{1}{2}}O\left(\tfrac{1}{N}\right)\} ,$$

where $\mathcal{X}_m$ is as defined in 4.1.2. Since $\frac{m}{n} \to 0$, $m^{\frac{1}{2}}O(N^{-1}) = O(N^{-1})$. Hence if c is a continuity point of the distribution of $f(W^\circ)$, then

$$\lim_{m,n\to\infty} \Pr\left\{ \left(\frac{n}{mN}\right)^{-\frac{1}{2}} d_{\mathcal{P}}(\mu_m,\gamma_N) \le c\right\}$$

$$= \lim_{m,n\to\infty} \Pr\left\{m^{\frac{1}{2}} d_{\mathcal{P}}(\mu_m,\gamma_N) \le c\right\} = \Pr\left\{f(W^\circ) \le c\right\} .$$

(iii) Let $\delta(m,n) = n^{\frac{1}{2}}/m$.

$$\Pr\{d_{\mathcal{P}}(\mu_m,\gamma_N) \le c\delta(m,n)\} = \Pr\{\mu_m(F) \le \gamma_N(F^{c\delta(m,n)}]) + c\delta(m,n) ; \text{ for all } F\in\mathcal{F}^{(2c\delta(m,n))}\} =$$

$$\Pr\{\mu_m(F) - \gamma_N(F) \le (\gamma_N(F^{c\delta(m,n)]}) - \mu(F^{c\delta(m,n)]}))$$

$$+ (\mu(F) - \gamma_N(F)) + \mu(F^{c\delta(m,n)]} - F) + c\delta(m,n) \; ;$$

$$\text{for all } F \in \mathcal{F}^{(2c\delta(m,n))}\}$$

$$= \Pr\{\tfrac{n}{m}(\gamma_N(F) - \nu_n(F)) \le q(F)\eta(F) + (\mu(F) - \gamma_N(F)) + c\delta(m,n)(q(F)+1);$$

$$\text{for all } F \in \mathcal{F}^{(2c\delta(m,n))}\} ,$$

where $-\frac{1}{N} \le \eta(F) \le \frac{1}{N}$,

$$= \Pr\{\tfrac{n}{m}(\mu(F) - \nu_n(F)) + \tfrac{n}{m}(\gamma_N(F) - \mu(F)) \le q(F)\eta(F) + (\mu(F) - \gamma_N(F))$$

$$+ c\delta(m,n)(q(F)+1); \text{ for all } F \in \mathcal{F}^{(2c\delta(m,n))}\}$$

$$= \Pr\{n^{\frac{1}{2}}(\mu(F) - \nu_n(F)) \le q(F)\eta(F)(2+ \tfrac{n}{m})\delta^{-1}(m,n) + c(q(F)+1) \; ;$$

$$\text{for all } F \in \mathcal{F}^{(2c\delta(m,n))}\}$$

$$= \Pr\{ \sup_{F \in \mathcal{F}^{(2c\delta(m,n))}} \frac{1}{q(F)+1} \int_F dX_n(t) \le \eta(F)(2+ \tfrac{n}{m})\delta(m,n)^{-1} + c\},$$

$\eta(F)(2 + \frac{n}{m})\delta(m,n)^{-1} = O(n^{\frac{1}{2}}/N) = o(1)$. Hence in the case $\frac{m}{n} \to \infty$,

$$\lim_{m,n\to\infty} \Pr\{ \left(\tfrac{n}{mN}\right)^{-\frac{1}{2}} d_{\mathcal{P}}(\mu_m, \gamma_N) \le c\}$$

$$= \lim_{m,n\to\infty} \Pr\{d_{\mathcal{P}}(\mu_m, \gamma_N) \le c \, \frac{n^{\frac{1}{2}}}{m}\} = \Pr\{f(W^\circ) \le c\} . \qquad \square$$

<u>Proposition 4.1.4</u>: $\left(\frac{N}{mn}\right)^{-\frac{1}{2}} d_{\mathcal{P}}(\mu_m, \nu_n) \Rightarrow f(W^\circ)$ as $m,n \to \infty$.

<u>Proof</u>: We treat separately the cases (i) $0 < a \le \frac{m}{n} \le b < \infty$ and

(ii) $\frac{m}{n} \to 0$. The case $\frac{m}{n} \to \infty$, i.e., $\frac{n}{m} \to 0$, follows due to the symmetric roles of m and n.

Let $\delta(m,n) = \left(\frac{N}{mn}\right)^{\frac{1}{2}}$.

(i)
$$\Pr\{d_{\mathcal{P}}(\mu_m,\nu_n) \le c\delta(m,n)\} = \Pr\{\mu_m(F) \le \nu_n(F^{c\delta(m,n)]})$$
$$+ c\delta(m,n);\ \text{for all}\ F \in \mathcal{F}^{(2c\delta(m,n))}\}$$
$$= \Pr\{\mu_m(F) - \nu_n(F) \le (\nu_n(F^{c\delta(m,n)]} - F) - \mu(F^{c\delta(m,n)]} - F))$$
$$+ \mu(F^{c\delta(m,n)]} - F) + c\delta(m,n);$$
$$\text{for all}\ F \in \mathcal{F}^{(2c\delta(m,n))}\}$$
$$= \Pr\{\delta(m,n)^{-1}(\mu_m(F) - \nu_n(F)) \le \delta(m,n)^{-1}(\nu_n(F^{c\delta(m,n)]} - F)$$
$$- \mu(F^{c\delta(m,n)]} - F)) + c(q(F)+1);$$
$$\text{for all}\ F \in \mathcal{F}^{(2c\delta(m,n))}\}$$
$$= \Pr\left\{\sup_{F \in \mathcal{F}^{(2c\delta(m,n))}} \frac{1}{q(F)+1} (Y_{m,n}(F) - Z_{m,n}(F)) \le c\right\} ,$$

where
$$Y_{m,n}(F) = \delta(m,n)^{-1}(\mu_m(F) - \nu_n(F))$$
and
$$Z_{m,n}(F) = \delta(m,n)^{-1}(\nu_n(F^{c\delta(m,n)]} - F) - \mu(F^{c\delta(m,n)]} - F)) .$$

$Y_{m,n}(t) = \delta(m,n)^{-1}(\mu_m[0,t] - \nu_n[0,t])$ is a random process on $D[0,1]$ and $Y_{m,n} \Rightarrow W^\circ$ as $m,n \to \infty$ where $0 < a \le \frac{m}{n} \le b < \infty$, by 1.4.2. Hence $f^{(c\delta(m,n))} \Rightarrow f(W^\circ)$. We can thus complete the proof for case (i) by showing that $\sup_{F \in \mathcal{F}^{(2c\delta(m,n))}} \frac{1}{q(F)+1} Z_{m,n}(F) \Rightarrow h(W^\circ)$, where $h(y) \equiv 0$

for $y \in D[0,1]$. For each $y \in D[0,1]$, let

$$h_{m,n}(y) = \left(\frac{N}{mn}\right)^{-\frac{1}{2}} \sup_{F \in \mathcal{F}^{(2c\delta(m,n))}} (q(F)+1)^{-1} \int_{F^{c\delta(m,n)}]_{-F}} dy .$$

It suffices to show that $h_{m,n}(\nu_n - \mu) \Rightarrow h(W^\circ)$. Since

$$\left| \left(\frac{N}{mn}\right)^{-\frac{1}{2}} (\nu_n - \mu) \right| \le (1+b) \left| n^{\frac{1}{2}} (\nu_n - \mu) \right| ,$$

and $n^{\frac{1}{2}}(\nu_n - \mu) \Rightarrow W^\circ$ by 1.2.5, it suffices (as in Lemma 4.1.1) to show that for any $y \in C[0,1]$ and $\{y_n\}$ in $D[0,1]$ such that $y_n \to y$,

$$\frac{1}{q(F)+1} \int_{F^{c\delta(m,n)}_{-F}} dy_n \to 0 \text{ , for each } F \in \mathcal{F}^{(2c\delta(m,n))} .$$

Let m and n be sufficiently large that $\sup_{0 \le t \le 1} |y_n[0,t] - y[0,t]| < \frac{\varepsilon}{3}$ and

$$\sup_{|s-t| < c\delta(m,n)} |y(s) - y(t)| < \frac{\varepsilon}{3} .$$

For any $F \in \mathcal{F}^{(2c\delta(m,n))}$,

$$\int_{F^{c\delta(m,n)}]_{-F}} dy \le q(F) \sup_{|s-t| < c\delta(m,n)} |y(s) - y(t)| < q(F)\frac{\varepsilon}{3} ,$$

$$\int_{F^{c\delta(m,n)}]_{-F}} dy_n - \int_{F^{c\delta(m,n)}]_{-F}} dy \le 2q(F) \sup_{0 \le t \le 1} |y_n[0,t] - y[0,t]| \le 2q(F)\frac{\varepsilon}{3},$$

as in the proof of 4.1.1. Hence

$$\frac{1}{q(F)+1} \int_{F^{c\delta(m,n)}]_{-F}} dy_n < \varepsilon .$$

(ii) $$\Pr\{d_{\mathcal{P}}(\mu_m,\nu_n) \le cm^{-\frac{1}{2}}\} = \Pr\{\mu_m(F) \le \nu_n(F^{cm^{-\frac{1}{2}}]}) + cm^{-\frac{1}{2}}, \text{ for all } F\in \mathcal{F}^{(2cm^{-\frac{1}{2}})}\}$$

$$= \Pr\{\mu_m(F) - \mu(F) \le (\nu_n(F^{cm^{-\frac{1}{2}}]}) - \mu(F^{cm^{-\frac{1}{2}}]})) + \mu(F^{cm^{-\frac{1}{2}}]}-F) + cm^{-\frac{1}{2}}\ ; \text{ for all } F\in\mathcal{F}^{(2cm^{-\frac{1}{2}})}\}$$

$$= \Pr\{X_m(F) - Z_{m,n}(F) \le (q(F)+1)c; \quad \text{for all } F\in\mathcal{F}^{(2cm^{-\frac{1}{2}})}\} ,$$

where

$$X_m(F) = m^{\frac{1}{2}}(\mu_m(F) - \mu(F)) \quad \text{and} \quad Z_{m,n}(F) = m^{\frac{1}{2}}(\nu_n(F^{cm^{-\frac{1}{2}}]}) - \mu(F^{cm^{-\frac{1}{2}}]})).$$

We wish to show that

$$\sup_{F\in\mathcal{F}^{(2cm^{-\frac{1}{2}})}} \frac{1}{q(F)+1} m^{\frac{1}{2}}(\nu_n(F^{cm^{-\frac{1}{2}}]}) - \mu(F^{cm^{-\frac{1}{2}}]}))$$

converges in probability to 0. But

$$\nu_n(F^{cm^{-\frac{1}{2}}]}) - \mu(F^{cm^{-\frac{1}{2}}]}) \le q(F) \sup_{0\le t\le 1} (\nu_n[0,t] - \mu[0,t]) ,$$

for any set $F\in\mathcal{F}^{(2cm^{-\frac{1}{2}})}$. So

$$\sup_{F\in\mathcal{F}} \frac{1}{q(F)+1} m^{\frac{1}{2}}(\nu_n(F^{cm^{-\frac{1}{2}}]}) - \mu(F^{cm^{-\frac{1}{2}}]}))$$

$$\le \left(\frac{m}{n}\right)^{\frac{1}{2}} (n^{\frac{1}{2}} \sup_{0\le t\le 1} (\nu_n[0,t] - \mu[0,t])) ,$$

which converges in probability to 0, since $\frac{m}{n} \to 0$. □

Proposition 4.1.5: For p = 1,2,... define $T_p(W^\circ)$ as $\sup_{0\le t_1\le\ldots\le t_p\le 1} |W^\circ(t_1)-W^\circ(t_2)+\ldots+(-1)^{p+1}W^\circ(t_p)|$. For $c > 0$,

$\{T_1(W^\circ) \le 2c,\ T_2(W^\circ) \le 3c,\ldots,T_{p-1}(W^\circ) \le pc,\ T_p(W^\circ) \le pc\} \subseteq \{f(W^\circ) \le c\} \subseteq$

$\{T_1(W°) \le 2c, T_2(W°) \le 3c, \ldots, T_{p-1}(W°) \le pc,\ T_p(W°) \le (p+1)c\}$.

Proof: $\{f(W°) \le c\} = \{T_1(W°) \le 2c,\ T_2(W°) \le 3c, \ldots\} \subseteq T_1(W°) \le 2c,$ $T_2(W°) \le 3c, \ldots, T_p \le (p+1)c\ \}$. Assume that $T_1 \le 2c$, $T_2 \le 3c, \ldots,$ $T_{p-1} \le pc$, $T_p \le pc$. Then $T_p \le (p+1)c$. We wish to show that this implies that $T_i \le (i+1)c$ for all i; assume for purposes of induction that for some $j > p$, $T_i \le (i+1)c$ for $i = 1,2,\ldots,j-1$.

By the triangle inequality, $T_j \le T_p + T_{j-p}$. Since $T_p \le pc$ and $T_{j-p} \le (j-p+1)c$, it follows that $T_j \le (j+1)c$. □

Using Proposition 1.4.3, we can evaluate $\Pr\{T_1(W°) \le 2c,\ T_2(W°) \le 2c\} = \Pr\{T_2(W°) \le 2c\}$, and $\Pr\{T_1(W°) \le 2c,\ T_2(W°) \le 3c\}$ directly to give rough upper and lower bounds on $\Pr\{f(W°) \le c\}$. The joint distribution of (T_1,T_2,T_3) seems intractable.

4.2 Asymptotic distribution for d_L.

Proposition 4.2.1:

a) $n^{\frac{1}{2}} d_L(\mu_n,\mu) \Rightarrow \sup_{0\le t\le 1} |\tfrac{1}{2}W°(T)|$.

b) $\left(\frac{n}{mN}\right)^{-\frac{1}{2}} d_L(\mu_m,\gamma_N) \Rightarrow \sup_{0\le t\le 1} |\tfrac{1}{2}W°(t)|$.

c) $\left(\frac{N}{mn}\right)^{-\frac{1}{2}} d_L(\mu_m,\nu_n) \Rightarrow \sup_{0\le t\le 1} |\tfrac{1}{2}W°(t)|$.

Proof: We shall prove only (a), since the proofs are completely analogous to 4.1.3, 4.1.4, and 4.1.5, respectively.

a) $\Pr\{d_L(\mu_n,\mu) \le cn^{-\frac{1}{2}}\} =$

$$\Pr\{\mu(-\infty, x-cn^{-\frac{1}{2}}) - cn^{-\frac{1}{2}} \leq \mu_n(x) \leq \mu(-\infty, x+cn^{-\frac{1}{2}}] + cn^{-\frac{1}{2}} ;$$

$$\text{for all } \; x \in [0,1]\}$$

$$= \Pr\{-2cn^{-\frac{1}{2}} \leq \mu_n(x) - \mu(x) \leq 2cn^{-\frac{1}{2}}; \;\; \text{for all } x \in [0,1]\}$$

$$= \Pr\{\sup_{0\leq t\leq 1} n^{\frac{1}{2}}|\mu_n(x) - \mu(x)| \leq 2c\} \to \Pr\{\sup_{0\leq t\leq 1} |W^\circ(t)| \leq 2c\} ,$$

by 1.4.2. □

Thus $d_L(\mu_n,\mu)$ has the same asymptotic distribution as $2d_K(\mu_n,\mu)$.

4.3 Absolute continuity of $f(W^\circ)$.

Definition 4.3.1: For $k \geq 1$ and $i=1,\ldots,k$, define

$$U_i = U_i^{(k)} = \sup_{\frac{i-1}{k} \leq t \leq \frac{i}{k}} W^\circ(t); \;\; V_i = V_i^{(k)} = \inf_{\frac{i-1}{k} \leq t \leq \frac{i}{k}} W^\circ(t) .$$

For $p \leq k$, define

$$T_p^{(k)}(W^\circ) = \max(\max_{1\leq i_1<\ldots<i_p\leq k} U_{i_1} - V_{i_2} + U_{i_3} - \ldots + (-1)^{p+1}U_{i_p} ,$$

$$\max_{1\leq i_1<\ldots<i_p\leq k} V_{i_1} - U_{i_2} + \ldots + (-1)^{p+1}V_{i_p}) \text{ if } p \text{ is odd,}$$

and

$$T_p^{(k)}(W^\circ) = \max(\max_{1\leq i_1<\ldots<i_p\leq k} U_{i_1} - V_{i_2} + U_{i_3} - \ldots + (-1)^{p+1}U_{i_p} ,$$

$$\max_{1\leq i_1<\ldots<i_p\leq k} V_{i_1} - U_{i_2} + \ldots + (-1)^{p+1}U_{i_p}) \text{ if } p \text{ is even.}$$

□

Remark 4.3.2: For a continuous sample path $W^\circ(t)$, $t\in[0,1]$, it is apparent that $T_p^{(k)}(W^\circ)$ may be defined in the same way as $T_p(W^\circ)$ with the restriction that no two of the points $t_1, t_2, \ldots, t_p$ (in the definition) are in the interval $[\frac{j}{k}, \frac{j+1}{k}]$, for any j. □

Definition 4.3.3: For $c > 0$, let $A(k,c)$ denote the event that

$$\max_{1 \le i \le k} \left(\sup_{\frac{i-1}{k} \le s \le t \le \frac{i}{k}} |W°(s) - W°(t)| \right) \le c. \qquad \square$$

We will establish the absolute continuity (with respect to Lebesgue measure) of $f(W°)$ by showing the following.

(I) $T_p^{(k)}(W°)$ is absolutely continuous for $1 \le p \le k$.

(II) For $p \ge 2$ if $T_p(W°) - T_{p-2}(W°) > c$ and $A(k,c)$ occurs then $T_p^{(k)}(W°) = T_p(W°)$, where $T_0 = 0$. Of course, $T_1^{(k)} = T_1$.

(III) As $c \downarrow 0$, $\Pr\{T_p(W°) - T_{p-2}(W°) > c\} \to 1$. For fixed c, as $k \to \infty$, $\Pr(A(k,c)) \to 1$.

(Note: It is not hard to show that $T_p^{(k)}(W°) \Rightarrow T_p(W°)$ as $k \to \infty$, in particular that $|T_p(W°) - T_p^{(k)}(W°)|$ is arbitrarily small with probability arbitrarily close to 1, for k sufficiently large. However, even when combined with (I), this weak convergence does not imply the absolute continuity of the distribution of $T_p(W°)$.)

(II) and (III) together imply

(IV) Given $\varepsilon > 0$, for k sufficiently large, $\Pr\{T_p^{(k)} = T_p\} > 1 - \varepsilon$. (Indeed, choose c first, then choose k.)

Proposition 4.3.4: (I) and (IV) together imply that $f(W°)$ is absolutely continuous.

Proof: We need the following lemmas:

Lemma 4.3.5: Let X be a random variable on some probability space. Let $X_1, X_2, \ldots$ be absolutely continuous random variables such that $\Pr\{X_k \neq X\} \to 0$ as $k \to \infty$. Then X is absolutely continuous.

Proof: Let A have Lebesgue measure zero. Then $\Pr\{X_k \in A\} = 0$, for all k. Choose $\varepsilon > 0$. $\Pr\{X \in A\} \leq \Pr\{X_k \in A\} + \Pr\{X_k \neq X\} < \varepsilon$, for k sufficiently large. Since ε is arbitrary, $\Pr\{X \in A\} = 0$.

□

It follows that if (I) and (IV) hold, then $T_p(W°)$ is absolutely continuous for each p.

Lemma 4.3.6: Let $X_1, X_2, \ldots$ be a sequence of random variables on some probability space and N a positive integer-valued random variable. If each random variable X_i is absolutely continuous, then X_N is absolutely continuous.

Proof: For any Lebesgue-measurable set A,

$$\Pr\{X_N \in A\} = \sum_{n=1}^{\infty} \Pr\{X_n \in A, N=n\} .$$

But if A has Lebesgue measure 0, $\Pr\{X_n \in A, N=n\} = 0$ for each n.

□

Lemma 4.3.7: $(p+1)^{-1} T_p(W°) \to 0$ a.s. as $p \to \infty$.

Before proving 4.3.7, we show that it will imply that $f(W°)$ is absolutely continuous, assuming (I) and (IV) are true. Indeed, by 4.3.7, with probability one, $f(W°) = \sup_{1 \leq p < \infty} (p+1)^{-1} T_p(W°) = (n+1)^{-1} T_n(W°)$ for some n, depending on $W°$.

Let $N = N(W°)$ be the least such value of n (define N in any manner on the probability zero set where there may not be such a value of n). Then by 4.3.5 and 4.3.6, $f(W°) = (N+1)^{-1}T_N(W°)$ is absolutely continuous. We complete the proof of 4.3.4 by proving 4.3.7.

Proof of 4.3.7: Let $M_k(W°) = \sup_{0 \le |s-t| \le 1/k} |W°(s)-W°(t)|$, $0 \le s, t \le 1$. Since $W°(t)$ is almost surely a uniformly continuous function on [0,1], $M_k(W°) \to 0$ a.s.

Let $0 \le t_1 < t_2 < \ldots < t_n \le 1$. The n points $t_1,\ldots,t_n$ can be partitioned into pairs and singletons such that for each pair, (t_i, t_{i+1}), both points t_i and t_{i+1} are in one of the intervals $[\frac{j}{k}, \frac{j+1}{k})$, $j=0,\ldots,k-2$, or $[\frac{k-1}{k}, 1]$, and such that there are at most k singletons. It easily follows that for any $n \ge k$

$$T_n \le T_k + \frac{n-k}{2} M_k(W°) .$$

Hence $(n+1)^{-1}T_n \le \frac{k}{n+1} T_1(W°) + \frac{1}{4} M_k(W°)$. Since $T_1(W°)$ is a.s. finite and $M_k(W°) \to 0$ a.s., it follows that $(n+1)^{-1}T_n(W°) \to 0$ a.s. as $n \to \infty$. □

We will prove (II) first and then (I) and (III).

Proposition 4.3.8: If $T_p(W°) - T_{p-2}(W°) > c$ and $A(k,c)$ occurs, then $T_p^{(k)}(W°) = T_p(W°)$ a.s.

Proof: Since $W°(t)$ is a.s. a uniformly continuous function on [0,1], assume that $W°(t)$ is uniformly continuous. Then we may choose points $0 \le t_1 \le \ldots \le t_p \le 1$ such that

$$T_p(W°) = (-1)^s(W°(t_1)-W°(t_2)+\ldots+(-1)^{p+1}W°(t_p)) \quad \text{for } s=0 \text{ or } 1.$$

Since by hypothesis, $T_p(W^\circ) - T_{p-2}(W^\circ) > c$, we must have $t_1<\ldots<t_p$. From the triangle inequality applied to the expression for $T_p(W^\circ)$, it follows that $T_p(W^\circ) \le T_{p-2}(W^\circ) + |W^\circ(t_i)-W^\circ(t_{i+1})|$ for $1 \le i \le p-1$. Thus if $T_p(W^\circ) - T_{p-2}(W^\circ) > c$, then $\min_{1\le i\le p-1} |W^\circ(t_i)-W^\circ(t_{i+1})| > c$. If in addition $A(k,c)$ occurs, this implies that for no values of j and i are $\frac{j-1}{k} \le t_i < t_{i+1} \le \frac{j}{k}$. Then by remark 4.3.2, $T_p^{(k)}(W^\circ) = T_p(W^\circ)$. □

<u>Proposition 4.3.9</u>: $T_p^{(k)}(W^\circ)$ is absolutely continuous for each $p \le k$.

<u>Proof</u>: Since the maximum of finitely many absolutely continuous random variables is itself absolutely continuous, it suffices to fix the values $1 \le i_1<\ldots<i_p \le k$ and (if p is even) show that $U_{i_1} - V_{i_2} + U_{i_3} - \ldots + (-1)^{p+1}V_{i_p}$ is absolutely continuous. (It will be seen that the same proof applies for $V_{i_1} - U_{i_2} + V_{i_3} - \ldots + (-1)^{p+1}U_{i_p}$, and for both cases when p is odd.)

Let $1 \le i_1<\ldots<i_p \le k$ be fixed for the remainder of the proof. We wish to consider the conditional joint distribution of $U_{i_1}, V_{i_2}, U_{i_3}, \ldots, V_{i_p}$ given $W^\circ(\frac{i_1-1}{k})$, $W^\circ(\frac{i_1}{k})$, $W^\circ(\frac{i_2-1}{k})$, $W^\circ(\frac{i_2}{k}),\ldots,$ $W^\circ(\frac{i_p-1}{k})$, $W^\circ(\frac{i_p}{k})$. The values $\frac{i_j}{k}$ and $\frac{i_{j+1}-1}{k}$ might not be distinct. Let $l_1 \ldots l_r$ be the points in order of the set $\{l:=\frac{i_j}{k} = l \text{ or } \frac{i_j-1}{k} = l,\ j=1,\ldots,p\}$, where $p \le r \le 2p$.

$Y(W^\circ) = (W^\circ(l_1),\ldots,W^\circ(l_r))$ has a (non-singular) r-variate normal distribution with density $\phi(X_1,\ldots,X_r)$. For any Lebesgue-measurable set A,

$$\Pr\{U_{i_1} - V_{i_2} + U_{i_3} - \ldots + (-1)^{p+1} V_p \in A\}$$

$$= \int \ldots \int \Pr\{U_{i_1} - V_{i_2} + U_{i_3} - \ldots + (-1)^{p+1} V_p \in A$$

$$/ W^\circ(l_1) = b_1, \ldots, W^\circ(l_r) = b_r\} \phi(b_1, \ldots, b_r)$$

$$db_1, \ldots, db_r \,.$$

We complete the proof by showing that if A has Lebesgue measure 0, $\Pr\{U_{i_1} - V_{i_2} + \ldots + (-1)^{p+1} V_{i_p} \in A / W^\circ(l_1) = b_1, \ldots, W^\circ(l_r) = b_r\} = 0$ for each $b_1, \ldots, b_r$.

Lemma 4.3.10: Let $0 \le c_1 < c_2 \le 1$ and let $Z(t) =$

$$W^\circ(c_1) + \frac{t-c_1}{c_2-c_1}[W^\circ(c_2) - W^\circ(c_1)] = -\frac{t-c_1}{c_2-c_1} W^\circ(c_2) + \frac{t-c_2}{c_2-t} W^\circ(c_1) \,.$$

Then for $c_1 < t < c_2$, $W^\circ(t) - Z(t)$ is a normal random variable and is independent of $W^\circ(s)$ for each $s \in [0,c_1] \cup [c_2,1]$.

(Note: this result bears out the intuitive notion that in the plane, the point $(t, E(W^\circ(t)/W^\circ(c_1), W^\circ(c_2)))$, for $c_1 < t < c_2$, lies on the line segment joining the points $(c_1, W^\circ(c_1))$ and $(c_2, W^\circ(c_2))$.)

Proof:

For $s \in [0,c_1]$ and $t \in (c_1,c_2)$, $E\, W^\circ(s) = 0 = E[W^\circ(t) - Z(t)]$, and it is easy to check (using the relationship $E\, W^\circ(t_1) W^\circ(t_2) = t_1(1-t_2)$ for $0 \le t_1 \le t_2 \le 1$) that $E\, W^\circ(s)(W^\circ(t) - Z(t)) = 0$. $W^\circ(t) - Z(t)$ is a linear combination of normal random variables and therefore normal, so $W^\circ(t) - Z(t)$ is independent of $W^\circ(s)$. Similarly for $s \in [c_1, 1]$. □

It follows that given $W°(l_1),\ldots,W°(l_r)$, $U_{i_1}, V_{i_2}, U_{i_3}, \ldots, V_{i_p}$ are mutually independent. If we can show that the conditional distribution of U_{i_1} is absolutely continuous, the absolute continuity of $T_p^{(k)}(W°)$ will follow.

Lemma 4.3.11: $\Pr\{\sup_{c_1 \le t \le c_2} W°(t) \le \lambda / W°(c_1) = b_1, W°(c_2) = b_2\} = 1 - e^{-2\lambda'(\frac{\lambda'-b}{c^2})}$, where $b = b_2 - b_1$, $\lambda' = \lambda - b_1$, $c = (c_2-c_1)^{\frac{1}{2}}$.

Proof: Conditional on the event $\{W°(c_1) = b_1$ and $W°(c_2) = b_2\}$, for $c_1 < t < c_2$, $W°(t)$ is the sum of the normal random variable $W°(t) - Z(t)$ and a constant $\frac{c_2-t}{c_2-c_1} b_2 - \frac{t-c_1}{c_2-c_1} b_1$.

It is easy to compute that for $c_1 \le s \le t \le c_2$ $E(W°(s) - Z(s))(W°(t) - Z(t)) = \frac{(s-c_1)(c_2-t)}{(c_2-c_1)}$. Making the change of variables $s' = \frac{s-c_1}{c_2-c_1}$ and $t' = \frac{t-c_1}{c_2-c_1}$, $0 \le s' \le t' \le 1$, and writing $X(s') = W°((c_2-c_1)s'+c_1) - Z((c_2-c_1)s'+c_1)$, we have $EX(s')X(t') = (c_2-c_1)s'(1-t')$, for $0 \le s' \le t' \le 1$. Hence $X(t')$, $0 \le t' \le 1$ in $(c_2-c_1)^{\frac{1}{2}}$ times a "Brownian Bridge" random process as defined in 1.4.1.

$$U_{i_1} = \sup_{c_1 \le t \le c_2} W°(t) - Z(t) = \sup_{0 \le t' \le 1} X(t') .$$

Hence

$$\Pr\{\sup_{c_1 \le t \le c_2} W°(t) \le \lambda / W°(c_1) = b_1, W°(c_2) = b_2\}$$

$$= \Pr\{\sup_{0\le t\le 1} (X(t') + Z((c_2-c_1)t'+c_1)) \le \lambda / W^\circ(c_1) = b_1,\ W'(c_2) = b_2\}$$

$$= \Pr\{\sup_{0\le t'\le 1} (X(t') + (b_2-b_1)t') + b_1 \le \lambda\}$$

$$= \Pr\{\sup_{0\le t'\le 1} (cW^\circ(t') + t'b) \le \lambda'\}, \text{ where } c = (c_2-c_1)^{\frac{1}{2}},\ b = b_2-b_1,\ \lambda' = \lambda - b_1,$$

$$= \Pr\{\sup_{0\le t<\infty} (cW^\circ(\tfrac{t}{t+1}) + \tfrac{t}{t+1} b) \le \lambda'\}$$

$$= \Pr\{\sup_{0\le t<\infty} (W(t) - \tfrac{\lambda'-b}{c} t - \tfrac{\lambda'}{c} \le 0\}, \text{ where } W \text{ is the Wiener process}$$

(see 1.4.1). By a result in [10], this last probability is equal to

$$1 - e^{-2\frac{\lambda'}{c}\cdot(\frac{\lambda'-b}{c})}. \qquad \square$$

We next prove III, in two parts.

Proposition 4.3.12. For each $c > 0$, $\Pr(A(k,c)) \to 1$ as $k \to \infty$, where $A(k,c)$ is as defined in 4.3.3.

Proof: We will consider $W^\circ(t) = W(t) - tW(1)$, $0 \le t \le 1$. Then

$$Z_i^\circ \underset{\text{def}}{=} \sup_{\frac{i-1}{k}\le s\le t\le\frac{i}{k}} |W^\circ(s) - W^\circ(t)|$$

$$= \sup_{\frac{i-1}{k}\le s\le\frac{i}{k}} (W^\circ(s) - W^\circ(\tfrac{i-1}{k})) - \inf_{\frac{i-1}{k}\le t\le\frac{i}{k}} (W^\circ(t) - W^\circ(\tfrac{i-1}{k}))$$

$$= \sup_{\frac{i-1}{k}\le s\le\frac{i}{k}} [(W(s)-W(\tfrac{i-1}{k})) + (s - \tfrac{i-1}{k})W(1)]$$

$$- \inf_{\frac{i-1}{k}\le t\le\frac{i}{k}} [(W(t)-W(\tfrac{i-1}{k})) + (t - \tfrac{i-1}{k})W(1)].$$

If $Z_i \underset{\text{def}}{=} \sup_{\frac{i-1}{k} \le s \le t \le \frac{i}{k}} |W(s)-W(t)|$, then it is clear that

$$|Z_i - Z_i^\circ| \le \frac{2}{k} W(1) \underset{p}{\to} 0 \quad \text{as} \quad k \to \infty .$$

Hence it will suffice to show that $\Pr\{\max_{1 \le i \le k} Z_i \le c\} \to 1$ as $k \to \infty$.

By the reflection principle, $\sup_{\frac{i-1}{k} \le s \le \frac{i}{k}} (W(s) - W(\frac{i-1}{k}))$ and $\inf_{\frac{i-1}{k} \le t \le \frac{i}{k}} (W(t) - W(\frac{i-1}{k}))$ are each distributed as twice a normal random variable with mean zero and variance $\frac{1}{k}$. Then

$$Z_i \le 2\left|\sup_{\frac{i-1}{k} \le s \le \frac{i}{k}} (W(s)-W(\tfrac{i-1}{k}))\right| + 2\left|\inf_{\frac{i-1}{k} \le t \le \frac{i}{k}} (W(t)-W(\tfrac{i-1}{k}))\right|$$

is stochastically smaller than 8 times the absolute value of a normal random variable with mean zero and variance $\frac{1}{k}$. Since $Z_1,\ldots,Z_k$ are mutually independent,

$$\Pr\{\max_{1 \le i \le k} Z_i \le c\} = (\Pr\{Z_i \le c\})^k \ge (\Phi(\frac{ck^{\frac{1}{2}}}{8}))^k .$$

We claim that $(\Phi(ck^{\frac{1}{2}}))^k \to 1$, which will complete the proof. Indeed, it is well known that

$$\Phi(ck^{\frac{1}{2}}) > 1 - \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{c^2k^{\frac{1}{2}}}{2}} (\frac{1}{ck^{\frac{1}{2}}}) ,$$

so

$$(\Phi(ck^{\frac{1}{2}}))^k > 1 - \frac{k^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} e^{-\frac{c^2k^{\frac{1}{2}}}{2}} ,$$

for k sufficiently large. Hence $(\Phi(ck^{\frac{1}{2}}))^k \to 1$ as $k \to \infty$. □

<u>Proposition 4.3.13</u>. For $p \geq 2$, $\Pr\{T_p(W^\circ) - T_{p-1}(W^\circ) > c\} \to 1$ as $c \downarrow 0$.

<u>Proof</u>: Let $A(K) = \{ \max_{1 \leq i \leq 2p} |W^\circ(\frac{i-1}{2p}) - W^\circ(\frac{i}{2p})| \leq K\}$. For fixed $\varepsilon > 0$, choose K sufficiently large that $\Pr(A) > 1 - \frac{\varepsilon}{2}$.

Let $0 \leq t_1 \leq \ldots \leq t_{p-2} \leq 1$ be chosen so that for $s = 0$ or $s = 1$, $T_{p-2}(W^\circ) = (-1)^s(W^\circ(t_1) - W^\circ(t_2) + \ldots + (-1)^{p-1}W^\circ(t_{p-2}))$; this can be done whenever $W^\circ(t)$ is a continuous function of t, i.e., almost surely.

For some i' and j', $1 \leq i' \leq p-1$ and $1 \leq j' \leq 2p$, $t_{i'-1} \leq \frac{j'-1}{2p} < \frac{j'}{2p} \leq t_{i'}$, where we define $t_0 = 0$, $t_{p-1} = 1$. Indeed, if $p \geq 2$, for some $i=1,\ldots,p-1$, $t_i - t_{i-1} \geq \frac{1}{p-1}$.

We wish to show that when A occurs, then for c sufficiently small, with probability greater than $1 - \frac{\varepsilon}{2}$ each interval $[\frac{j-1}{2p}, \frac{j}{2p}]$, $j=1,\ldots,2p$, contains points u and v such that $u < v$ and $W^\circ(u) - W^\circ(v) > c$ and points u' and v' such that $u' < v'$ and $W^\circ(v') - W^\circ(u') > c$. Since such points therefore can be found in the interval $[\frac{j'-1}{2p}, \frac{j'}{2p}]$ defined in the last paragraph, then either $(-1)^s[(-1)^{i'}(W^\circ(u) - W^\circ(v))] > c$ or $(-1)^s[(-1)^{i'}(W^\circ(u') - W^\circ(v'))] > c$, so $T_p(W^\circ) - T_{p-2}(W^\circ) > c$.

Assume for definiteness that $W^\circ(\frac{j'-1}{2p}) = b_1$ and $W^\circ(\frac{j'}{2p}) = b_2$ with $b_1 \leq b_2$. As in 4.3.11, for $\lambda \geq b_2$,

$$\Pr\{ \sup_{\frac{j'-1}{2p} \leq t \leq \frac{j'}{2p}} W^\circ(t) \leq \lambda / W^\circ(\tfrac{j'-1}{2p}) = b_1, W^\circ(\tfrac{j'}{2p}) = b_2\} = 1 - e^{-4p(\lambda - b_2)(\lambda - b_1)} .$$

Under the assumptions $W^\circ(\frac{j'-1}{2p}) = b_1 \le b_2 = W^\circ(\frac{j'}{2p})$, if $\sup_{\frac{j'-1}{2p}\le t\le\frac{j'}{2p}} W^\circ(t) > b_2 + c$, then points u and v , u' and v' as described may be found; indeed in the interval $[\frac{j'-1}{2p}, \frac{j'}{2p}]$, the process starts at b_1, goes up to b_2+c and back down to b_2. Then

$$\Pr\{\sup_{\frac{j'-1}{2p}\le t\le\frac{j'}{2p}} W^\circ(t) \le b_2 + c/W^\circ(\tfrac{j'-1}{2p}) = b_1,\ W^\circ(\tfrac{j'}{2p}) = b_2\}$$

$$= 1 - e^{-4p\ c(c+b_2-b_1)} \le 1 - e^{-4p\ c(c+K)} ,$$

under the assumption that A(K) occurs. If c is chosen sufficiently small, this value may be made smaller than $\frac{\varepsilon}{2}$. □

We have now proved (I), (II), and (III), showing by 4.3.4 that $f(W^\circ)$ is absolutely continuous.

CHAPTER 5

Statistical Properties

5.1 Related test statistics.

From among the metrics which generate the topology of weak convergence on $\mathcal{P}(S)$, we have elected to examine d_P and d_L because they are the ones which have appeared most often in the literature. As an alternative, we may consider, for any $b > 0$, $d_P^{(b)}(P,Q) = \inf\{\varepsilon>0: P(F) \le Q(F^\varepsilon) + b\varepsilon\ ;\ \text{all closed } F\subseteq S\}$ or $d_L^{(b)}(P,Q) = \inf\{\varepsilon>0: P(-\infty,x-\varepsilon] - b\varepsilon \le Q(-\infty,x] \le P(-\infty,x+\varepsilon] + b\varepsilon\ ;\ \text{for all } x \in E^k\}$, which are topologically equivalent to d_P and d_L and have respectively the same uniformities. (Indeed, if $b > 1$, $d_P^{(b)}(P,Q)/b \le d_P(P,Q) \le bd_P^{(b)}(P,Q)$, and $d_L^{(b)}(P,Q)/b \le d_L(P,Q) \le bd_L(P,Q)$; if $b < 1$, reverse the inequalities.)

The two-sample test (for equal sample sizes) based on $d_P^{(2)}(\mu_n,\nu_n)$ seems to have advantages over that based on $d_P(\mu_n,\nu_n)$ because of differences in the null distribution for finite sample sizes. $\Pr\{d_P(\mu_n,\nu_n) \le c\}$ is constant in the intervals $[i/2n,\ (i+1)/2n)$, $1 \le i \le 2n$, because $\mu_n(F) - \nu_n(F^c)$ is always a multiple of $1/n$, and the points at which the atoms of μ_n and ν_n are located are separated by multiples of $1/2n$. Thus, intuitively, there is more "going on" at the points $2/2n, 4/2n,\ldots,1$ than at the points $1/2n, 3/2n,\ldots,(2n-1)/2n$.

For $d_P^{(2)}(\mu_n,\nu_n)$ the interval between atoms is the same as the common divisor of the possible values of the test statistic and so there is no such distinction between points of the form $2i/2n$ and those of the form $(2i+1)/2n$. This difference between d_P and $d_P^{(2)}$ can be seen in the tables in the appendix. The result seems to be a faster, or at least more uniform convergence of $d_P^{(2)}(\mu_n,\nu_n)$ to its asymptotic distribution. A further difference is that given our restriction on the values of z, as described in Chapter 3, we can construct more extensive tables of the distribution of $d_P^{(2)}$, using the same methods as for d_P. Indeed, in place of 3.1.1, we have

$$\Pr\{d_P^{(2)}(\mu_n,\nu_n) \le c\} = \binom{2n}{n}^{-1} \sum_{0\le p\le z} A(p,n,z) \quad .$$

The rate of convergence to the asymptotic distribution for the one-sample d_P test statistic seems roughly on the order of that of the Kolmogorov test judging from empirical distribution functions based on actual randomly-generated samples. Our exact distributions show that the convergence for the two-sample d_P test is slower than for the two-sample Kolmogorov-Smirnov test. The proof of Proposition 4.1.1 may give some explanation for this, but we have not been able to give a precise explanation.

The Monte Carlo one-sample distribution functions indicate that the upper bound given by $\Pr\{T_1(W^\circ) \le 2c,\ T_2(W^\circ) \le 3c\}$ is fairly close to the true asymptotic distribution and the lower bound, $\Pr\{T_2(W^\circ) \le 2c\}$ is not as good.

In [26], Rosenblatt suggests a test based on the metric $d_2(P,Q) = \sup_{I \text{ an interval}} |P(I) - Q(I)|$, stating that it has somewhat better power than d_K against certain alternatives. The asymptotic distribution for this test statistic, multiplied by $n^{\frac{1}{2}}$, is $\Pr\{T_2(W°) \leq c\}$. As a generalization of d_2, we suggest:

(1) $d_k(P,Q) = \sup|P(F) - Q(F)|$, the supremum being taken over all $F \in \bigcup_{i=1}^{k} \mathcal{F}_i$, where $\mathcal{F}_i = \{F \in \mathcal{F}: q(F) = i\}$, $q(F)$ being defined as in Chapter 4, and

(2) $d_P^*(P,Q) = \inf\{\varepsilon > 0:$ for $k=1,2,\ldots,$ $|P(F) - Q(F)| \leq (k+1)\varepsilon$; all $F \in \mathcal{F}\}$.

d_P^* is thus the same as d_P except that the "end effects" for each interval are ignored; d_P^* and d_P put a "penalty" of a factor of $(k+1)$ on the number of component intervals (more precisely on $q(F)$) allowed for the set F under consideration. Some such penalty is necessary if the metric is to take into account all $F \in \mathcal{F}$; indeed, $\Pr\{T_k(W°) \leq c\} \to 0$ as $k \to \infty$ (because $W°(t)$ a.s. has infinite total variation). Intuition does not indicate why the factor should be $k+1$. However, a factor of k would not be appropriate, since by the triangle inequality, $|P(F) - Q(F)| \leq \varepsilon$ for all $F \in \mathcal{F}_1$ implies that for $k=1,2,\ldots,$ $|P(F) - Q(F)| \leq k\varepsilon$ for all $F \in \mathcal{F}_k$.

It should be noted that d_P^* does not generate the same topology on $\mathcal{P}(S)$ as d_P and d_L. However, we would expect that the power properties of d_P (relative to d_K) which will be described in the next section, would be shared by d_P^* .

5.2 Power of the test based on d_P.

We first show that the tests based on d_P and d_L are consistent. Let U be the uniform measure on $[0,1]$. Let $V \in \mathcal{P}([0,1])$, $V \neq U$, and let V_n be the empirical measure of a sample of size n from V. We test the hypothesis that $V = U$ by using one of the test statistics $d_L(V_n,U)$ or $d_P(V_n,U)$. By Proposition 1.3.5, $d_P(V_n,U) \to d_P(V,U) > 0$ a.s. and $d_L(V_n,U) \to d_L(V,U)$ a.s. Since, if F is a continuous distribution function, $\{F(X_i)\}$ will be a sample from U if and only if the sample $\{X_i\}$ actually comes from F, our one-sample test is consistent against all alternatives to a continuous null distribution.

To show consistency for the two-sample test, we use the following result related to Proposition 1.3.12.

Proposition 5.2.1: Let P_n put mass $1/n$ on n of the points $\{1/2n, 2/2n,\ldots,1\}$ and let Q_n put mass $1/n$ on the other n points. Then $d_P(P_n,Q_n) \geq d_K(P_n,Q_n)/3 - 1/n$.

Proof: If for some $x \in [0,1]$, $|P_n[0,x] - Q_n[0,x]| \geq 3\varepsilon > 0$, then either

$$P_n([0,x]) \geq Q_n([0,x]^\varepsilon) + \varepsilon - 1/n \text{ , or}$$
$$P_n([0,x]) \geq Q_n([x,1]^\varepsilon) + \varepsilon - 1/n \text{ .}$$

Indeed, $Q_n([x,x+\varepsilon]) \leq (2n\varepsilon+1)/n$, and $Q_n([x-\varepsilon,x]) \leq (2n\varepsilon+1)/n$. Hence

$$d_P(P_n,Q_n) \geq \varepsilon - 1/n \text{ .} \qquad \square$$

(By using a more laborious argument along the lines of 4.1.5, $d_K(P_n,Q_n)/3$ can be replaced by $d_K(P_n,Q_n)/2$.) The consistency of the d_P and d_L two-sample tests now follows from the consistency of the Kolmogorov-Smirnov test.

Massey [20] has given an example to show that the Kolmogorov-Smirnov test is not unbiased. Similarly, we can show that the one-sample test based on d_P is not unbiased. Suppose we take a sample $\{X_i\}$ of size n from the probability measure V which equals Lebesgue measure on $(c_\alpha n^{-\frac{1}{2}}, 1-c_\alpha n^{-\frac{1}{2}})$, where c_α is the critical value for the size α test, and which places mass $c_\alpha n^{-\frac{1}{2}}$ at the points $c_\alpha n^{-\frac{1}{2}}$ and $1 - c_\alpha n^{-\frac{1}{2}}$. (Assume that n is sufficiently large that $c_\alpha n^{-\frac{1}{2}} < 1 - c_\alpha n^{-\frac{1}{2}}$.) We may generate such a sample $\{X_i\}$ by "trimming" a sample $\{X_i'\}$ from U[0,1], i.e., by letting $X_i = X_1'$ if $X_i' \in (c_\alpha n^{-\frac{1}{2}}, 1 - c_\alpha n^{-\frac{1}{2}})$, and letting $X_i = c_\alpha n^{-\frac{1}{2}}$ if $X_i' \le c_\alpha n^{-\frac{1}{2}}$, and $X_i = 1 - c_\alpha n^{-\frac{1}{2}}$ if $X_i' \ge 1 - c_\alpha n^{-\frac{1}{2}}$. Let U_n^* be the empirical measure of such a trimmed sample from U and let U_n the the empirical measure of the original sample from U.

For any sample $\{X_i'\}$ from U, $d_P(U_n^*,U) \ge c_\alpha n^{-\frac{1}{2}}$ implies that $d_P(U_n,U) \ge c_\alpha n^{-\frac{1}{2}}$. Indeed, for any closed set $F \subseteq [0,1]$,

$$U_n^*(F^{c_\alpha n^{\frac{1}{2}}]}) \ge U_n(F^{c_\alpha n^{-\frac{1}{2}}]}) ,$$

so

$$\sup_{\mathit{closed}\ F\subseteq[0,1]} U(F) - U_n(F^{c_\alpha n^{-\frac{1}{2}}]}) \ge \sup_{\mathit{closed}\ F\subseteq[0,1]} U(F) - U_n^*(F^{c_\alpha n^{-\frac{1}{2}}]}).$$

With positive probability, $d_P(U_n,U) > c_\alpha n^{-\frac{1}{2}}$ but $d_P(U_n^*,U) < c_\alpha n^{-\frac{1}{2}}$. For example, if U_n has about $n - 4c_\alpha n^{+\frac{1}{2}}$ points close to being uniformly distributed on the interior of $(2c_\alpha n^{-\frac{1}{2}}, 1-2c_\alpha n^{-\frac{1}{2}})$ and the remainder evenly divided between small intervals about 0 and 1, then $d_P(U_n,U)$ is close to $\frac{4}{3} c_\alpha n^{-\frac{1}{2}}$ (worst set is the union of the small intervals about 0 and 1) and $d_P(U_n^*, U)$ is close to $\frac{4}{5} c_\alpha n^{-\frac{1}{2}}$ (worst set $\{c_\alpha n^{-\frac{1}{2}}, 1-c_\alpha n^{-\frac{1}{2}}\}$). Hence the one-sample test based on d_P is not unbiased.

With additional complications, a similar example can be constructed for the two-sample case.

To attempt to gain some insight into the power of the d_P test against various kinds of alternatives we have made some empirical comparisons of power of the one- and two-sample tests based on d_P with that of the test of the same size based on d_K. For the one-sample test the d_P critical value was taken from the distribution functions which were estimated in a separate Monte Carlo simulation. For the sake of convenience, for the two sample tests the size was chosen so that a non-randomized test could be used for both d_K and d_P. Random samples were generated from several alternatives to the null distribution. The results of these comparisons are given in the following table.

One-sample tests $\alpha = .05$

Description	(sample size)	Number Trials	Number Rejected d_K	Number Rejected d_P	Proportion Rejected d_K	Proportion Rejected d_P
U[0,1] vs. U[0,1]	(40)	4000	189	194	.047	.048
U[0,1] vs. U^2	(40)	2000	1899	1900	.950	.950
U[0,1] vs. 2-sided U^2	(40)	2000	743	1445	.371	.722
U[0,1] vs. bimodal	(40)	2000	287	358	.143	.179
U[0,1] vs. trimodal	(40)	2000	173	195	.086	.097
U[0,1] vs. U[0,1]	(100)	1000	47	53	.047	.053
U[0,1] vs. 2-sided U^2	(100)	1000	923	999	.923	.999
U[0,1] vs. bimodal	(100)	1000	293	672	.293	.672
U[0,1] vs. trimodal	(100)	1000	150	239	.150	.239

Two-sample tests $\alpha = .25$

Description	(sample size)	Number Trials	Number Rejected d_K	Number Rejected d_P	Proportion Rejected d_K	Proportion Rejected d_P
N(0,1) vs. N(0,1)	(30)	200	46	48	.230	.240
N(0,1) vs. N(.5,1)	(30)	1000	703	694	.703	.694
N(0,1) vs. N(.75,1)	(30)	1000	914	913	.914	.913
N(0,1) vs. N(1,1)	(30)	1000	987	983	.987	.983
N(0,1) vs. N(0,2)	(30)	1000	369	438	.369	.438
N(0,1) vs. N(0,4)	(30)	1000	693	825	.693	.825
U[0,1] vs. U^2	(30)	1000	808	798	.808	.798
U[0,1] vs. U^3	(30)	1000	988	988	.988	.988
U[0,1] vs. 2-sided U^2	(30)	1000	562	657	.562	.657

Explanation: U[0,1] denotes the uniform distribution on [0,1]. $N(\mu,\sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. U^k is the "Lehmann" alternative generated as $\max(U_1,\ldots,U_k)$, where $U_1,\ldots,U_k$

are uniform random variables. "2-sided U^2", "bimodal" and "trimodal" are random variables with respectively unimodal, bimodal, and trimodal distribution functions generated by taking appropriate mixtures of random variables with distribution U^2; specifically if B_1 and B_2 are independent Bernoulli random variables mutually independent of V, a random variable with distribution U^2, then $V_1 = B_1V/2 + (1-B_1)(1-V/2)$ is our "2-sided" random variable, and $V_2 = B_2V_1/2 + (1-B_2)(.5+V_1/2)$ is our "bimodal" random variable. The "trimodal" random deviates were generated as follows. Let B take the values 0,1, and 2 with equal probability. Then $V_3 = \frac{B}{3} + V_1/3$. Only uniform random numbers were used in the one-sample tests, because the time required to evaluate $\Phi(X_i)$ for the normal distribution function Φ threatened to become excessive for the large samples and large number of trials involved. The "2-sided", "bimodal" and "trimodal" alternatives were suggested by the examples given below, as being potentially favorable to d_P, because d_P can "look at" several intervals simultaneously, not just at one as does d_2. The evidence is not sufficient to substantiate or refute such a conjecture.

It can be seen that in general d_P performed better than d_K, except in the case of alternatives which were stochastically greater than the null distribution, in which case the differences are not significant.

Rosenblatt [26] gives examples comparing d_2 and d_K which are useful for our purposes, shedding some light on the above results. An example in which d_2 performs better is an alternative V to U[0,1] which has distribution function

$$F(x) = .03, \quad 0 \le x \le .06$$

$$F(x) = x, \quad x \ge .06 ,$$

where the sample size $n = 1600$ and $\alpha = .05$. In this case the acceptance region for the d_K test is (using [29]) $d_K(V_n,U) > 1.36/40 = .034$, and for d_2 is (from Rosenblatt's results) $d_2(V_n,U) > 1.75/40 = .044$. A sample from V will always satisfy $d_2(V_n,U) \ge .06$, since there are no observations in $(0,.06)$, and hence the d_2 test rejects with probability 1. On the other hand, the d_K test has power close to .05, since it can be shown (for example as Rosenblatt suggests, using a result of Uspensky cited in [25]) that $\Pr\{\sup_{0 \le t \le .06} |U[0,t] - V_n[0,t]| > .034\}$ is small for this value of n.

The dramatic difference between d_2 and d_K in this case is due to the proper choice of n and, of course, V. d_P is also similarly superior to d_2 in this case; the critical value for the d_P test is at most $.7/40 = .0175$, based on our Monte Carlo tables, but $d_P(V_n,U) \ge .02$ (consider the set $(0,.06)$), so the d_P test rejects with probability 1. However, if $F = .0225$ on $[0,.045]$, then d_P also has poor power, while d_2 rejects a.s. (Rosenblatt also gives an example,

$$\begin{aligned} F(x) &= x; \quad 0 \le x < .5 - \delta \\ &= .5 - \delta; \quad .5 - \delta \le x \le .5 + \delta \\ &= x; \quad .5 + \delta < x \le 1, \end{aligned}$$

for which d_K is superior to d_2 and also to d_P, for proper choice of δ and n.

As an example of the same type in which d_P performs better than either d_2 or d_K, let $n = 1600$, $\alpha = .05$, and let V place mass .01 at the points 0 and .04 and mass .02 at the point .02, and be equal to Lebesgue measure on (.04,1]. Then if V_n is the empirical measure of a sample of size n from V, $d_P(V_n,U) \geq .02$ (consider the set {0,.02}), so d_P rejects with probability 1. However, $d_K(V,U) = .02 = d_2(V,U)$, while their critical values are .034 and .044 respectively, so as in the previous examples, it can be shown that the d_K and d_2 tests have low power for the values of n and α chosen.

Massey [20] gives a lower bound for the power of the Kolmogorov-Smirnov test against a class of alternatives to the null distribution P of the form $\{P' \in \mathcal{P}(S): d_K(P,P') = c\}$. The result provides an alternative proof of consistency of the test. We give a similar result for d_P.

<u>Proposition 5.3.2</u>: Let P and Q be elements of $\mathcal{P}(S)$ and P_n and Q_n be the empirical measures of samples of size n from P and Q respectively. Let $\Pr\{d_P(P_n,P) \geq c_\alpha\} = \alpha$. If $Q \in \{P' \in \mathcal{P}(S): d_P(P,P') = c\}$ for some $c > c_\alpha$, then $\Pr\{d_P(Q_n,P) \geq c_\alpha\} \geq 1 - \Pr\{Y - \frac{n}{2} \leq -n(c-c_\alpha)\}$, where Y is a $B(n,\frac{1}{2})$ (i.e., binomial based on n trials with probability $\frac{1}{2}$ of success on each trial) random variable.

<u>Proof</u>: Since $d_P(Q,P) = c$, there exists a closed set $F_0 \subseteq S$ such that $Q(F_0) \geq P(F_0^{c]}) + c$.

$$\begin{aligned}
&\Pr\{d_P(Q_n,P) > c_\alpha\} \\
&\quad = 1 - \Pr\{Q_n(F) \leq P(F^{c_\alpha]}) + c_\alpha \text{ ; all closed } F \subseteq S\} \\
&\quad \geq 1 - \Pr\{Q_n(F_0) \leq P(F_0^{c_\alpha]}) + c_\alpha\}
\end{aligned}$$

$$\geq 1 - \Pr\{Q_n(F_0) \leq Q(F_0) - c + c_\alpha\}$$

$$= 1 - \Pr\{n(Q_n(F_0) - Q(F_0)) \leq -n(c-c_\alpha)\} \ .$$

$nQ_n(F_0)$ is a $B(n,Q(F_0))$ random variable. $\Pr\{n(Q_n(F_0) - Q(F_0)) \leq -n(c-c_\alpha)\}$ is maximized for $Q(F_0) = \frac{1}{2}$, since the binomial random variable has maximum variance in this case. □

APPENDIX

A.1 Tables of the one-sample distribution function.

The distribution function of $n^{\frac{1}{2}}d_p(\mu_n,\mu)$ has been estimated for $n = 5,10,20,40,60$ and 80, where μ is uniform measure on $[0,1]$ and μ_n is the empirical measure of a sample of size n from μ. Our procedure was to generate 4000 random samples of size n, to compute for each sample the value of $d_p(\mu_n,\mu)$ and record the number of samples for which the value fell in the interval $[.001i, .001(i+1))$ for $i=0,\ldots,199$ and $[.2,\infty)$. (For $n = 5$ and $n = 10$, we used $[.002i, .002(i+1))$.) To get the tables which give $\Pr\{n^{\frac{1}{2}}d_p(\mu_n,\mu) \le c\}$ for values of c which are multiples of $.01$, we interpolated linearly between the values estimated for $\Pr\{n^{\frac{1}{2}}d_p(\mu_n,\mu) \le .001i\cdot n^{\frac{1}{2}}\}$ and $\Pr\{n^{\frac{1}{2}}d_p(\mu_n,\mu) \le .001(i+1)n^{\frac{1}{2}}\}$ for $i = 0,\ldots,199$.

The random samples from μ were generated using a program from the UNC Computation Center's "Scientific Subroutine Package." The program uses the "congruential" method, as explained in [Maclaren and Marsaglia, *J.A.C.M. 12*, 83-84]. It is noted in this article that the random numbers produced by such a method may not be mutually independent, and a correction suggested in the article has been used: 100 numbers generated by this method are placed in a table and labeled 00,...,99; two digits from an additional random number are used to determine which location in the table to take the next "observation" from, and the observation used is replaced by a new random number.

The same procedure was used to generate the U[0,1] random sample for the power comparisons of Chapter 5. The Box-Mueller method (see Bell *C.A.C.M. 11*, 498 and Knop *C.A.C.M. 12*, 281) was used to generate the normally distributed random numbers for the power comparisons.

Several computer programs were written to compute d_p using the methods of Chapter 2: one to use the Hungarian algorithm in the general two-sample case; one for the case of two discrete measures on E^1, and one for the one-sample test statistic on E^1. The programs have been checked by using them to find d_p in a number of cases which can be computed by hand, involving various arrangements of the points of the "worst set." The two-sample programs provide a check on each other.

The programs for E^1 take on the order of 75 seconds to perform 4000 evaluations of $d_p(\mu_n,\mu)$ for n = 80. The program written for the Hungarian algorithm is considerably slower, taking on the order of 30 seconds to perform a single evaluation for n = 30, and the time required increases rapidly with n. The Hopcroft-Karp algorithm presumably will permit the evaluation for general metric spaces for larger values of n.

One-sample distribution: $Pr\{\sqrt{n}\ d_p(\mu_n,\mu) \le x\}$

x	n=5	10	20	40	60	80
.21	.000	.000	.000	.000	.000	.000
.22	.000	.001	.001	.000	.000	.000
.23	.002	.002	.003	.001	.000	.001
.24	.007	.005	.004	.003	.002	.001
.25	.012	.010	.008	.006	.003	.002
.26	.021	.018	.015	.011	.008	.005
.27	.035	.028	.022	.015	.012	.012
.28	.052	.038	.033	.021	.017	.017
.29	.070	.056	.052	.031	.026	.029
.30	.095	.081	.069	.054	.037	.046
.31	.124	.103	.096	.071	.063	.060
.32	.156	.131	.120	.098	.083	.078
.33	.194	.158	.145	.119	.103	.109
.34	.236	.194	.184	.140	.141	.131
.35	.279	.235	.219	.180	.165	.157
.36	.322	.282	.262	.208	.199	.209
.37	.358	.318	.301	.253	.236	.240
.38	.400	.356	.336	.286	.264	.271
.39	.445	.392	.384	.332	.307	.319
.40	.487	.440	.421	.368	.360	.362
.41	.527	.476	.457	.416	.392	.392
.42	.560	.513	.501	.448	.431	.444
.43	.592	.552	.530	.481	.482	.472
.44	.628	.592	.568	.526	.509	.501
.45	.655	.627	.604	.555	.561	.541
.46	.684	.660	.630	.598	.589	.590
.47	.709	.688	.659	.622	.613	.618
.48	.731	.712	.687	.660	.655	.641
.49	.753	.738	.712	.685	.676	.678
.50	.774	.764	.739	.706	.710	.701
.51	.794	.784	.760	.737	.736	.720
.52	.815	.804	.784	.756	.754	.758
.53	.833	.828	.801	.776	.776	.770
.54	.848	.841	.818	.795	.804	.785
.55	.863	.855	.837	.820	.818	.807
.56	.877	.869	.850	.834	.843	.828
.57	.889	.881	.863	.855	.854	.842
.58	.901	.894	.877	.867	.866	.863
.59	.913	.905	.888	.878	.884	.877
.60	.923	.915	.899	.893	.892	.884

x	n=5	10	20	40	60	80
.61	.933	.922	.907	.902	.899	.894
.62	.941	.930	.915	.912	.911	.909
.63	.947	.938	.925	.919	.920	.915
.64	.952	.944	.932	.927	.928	.926
.65	.958	.951	.940	.934	.935	.936
.66	.962	.958	.950	.941	.941	.942
.67	.967	.963	.954	.950	.948	.947
.68	.970	.967	.959	.956	.955	.954
.69	.973	.970	.963	.959	.960	.960
.70	.974	.973	.968	.963	.966	.964
.71	.978	.978	.972	.968	.968	.969
.72	.981	.980	.974	.970	.971	.973
.73	.983	.983	.976	.975	.974	.975
.74	.986	.984	.978	.978	.977	.979
.75	.988	.985	.980	.982	.980	.982
.76	.990	.988	.983	.984	.982	.984
.77	.991	.990	.986	.986	.983	.985
.78	.992	.991	.988	.988	.986	.987
.79	.994	.992	.989	.988	.987	.988
.80	.995	.994	.990	.989	.989	.990
.81	.996	.995	.992	.990	.990	.992
.82	.996	.996	.993	.991	.991	.992
.83	.996	.996	.994	.992	.993	.994
.84	.997	.997	.995	.993	.993	.995
.85	.998	.997	.996	.994	.994	.996
.86	.999	.997	.996	.994	.995	.996
.87	.999	.998	.997	.995	.996	.996
.88	.999	.999	.997	.996	.997	.997
.89	.999	.999	.998	.996	.997	.997
.90	.999	.999	.999	.996	.998	.998
.91	.999	.999	.999	.996	.998	.998
.92	.999	.999	.999	.997	.998	.999
.93	.999	.999	.999	.997	.998	.999
.94	.999	.999	.999	.998	.998	.999
.95	.999	.999	.999	.998	.999	.999

A.2 Tables of the two-sample distribution function.

The values of $K(i,n,p,s,t)$ were computed for $i = 0,1,2$; $z = 0,1,\ldots,12$; and $n = 1,2,\ldots,100$, for each value of d,s, and t, $0 \le p \le z$, $0 \le s \le \min(z,p)$, using equation 3.5.2 with the simplifications described in Section 3.6. $A(p,n,z)$ as defined in 3.1.1 is then computed as $\sum_{i=0}^{2} \sum_{s=0}^{z} \sum_{t=0}^{p} K(i,n,p,s,t)$, where as described in Section 3.1, $K(i,n,p,s,t)$ depends on z.

In the following tables we have listed $\binom{2n}{n}^{-1} \sum_{0 \le p \le z/2} A(p,n,z)$, which is $\Pr\{d_p(\mu_n,\nu_n) \le z/2n\}$, and also $\binom{2n}{n}^{-1} \sum_{0 \le p \le z} A(p,n,z)$, which is $\Pr\{d_p^{(2)}(\mu_n,\nu_n) \le z/2n\}$, where $d_p^{(2)}$ is defined as in Section 5.1.

Because of the recursiveness in equations 3.5.2, there is reason to suspect that rounding errors will affect the results. To check for this, the computations were repeated with values of n up to 50 for larger values of z, using "double precision" storage in the computer memory. Storing the $K(\cdot,\cdot,\cdot,\cdot,\cdot)$ values in double precision did not affect the result while storing the $D(\cdot,\cdot)$ values in double precision made a difference in the fourth decimal place of the computed probabilities. (There was no marked effect for smaller values of z.) The tables presented here were computed with the $D_i(\cdot,\cdot)$ values in double precision Storage limitations preclude the use of double precision for all the numbers for the full range of parameters.

The $D_i(l,s)$ values can be checked by hand for moderately small values of ℓ, s and z, sufficient to test the operation of all parts of this segment of the program. The probabilities in the table have been checked for n = 6, 7 and 8 for several values of z, by generating all permutations of the two samples and using the two-sample program for E^1 to find $d_p(\mu_n, \nu_n)$ for each permutation. An additional check on the correctness of the programming (although not necessarily of the theory) is that for z large enough $\sum_{p=0}^{z} A(p,n,z) = \binom{2n}{n}$.

$\Pr\{d_P(\mu_n,\nu_n) \le z/2n\}$

n	z=1	2	3	4	5	6	7	8	9	10	11	12
1	1.0000											
2	.6667	1.0000										
3	.4000	1.0000										
4	.2286	.9429	1.0000									
5	.1270	.8651	.9683	1.0000								
6	.0693	.7705	.9199	1.0000								
7	.0373	.6719	.8631	.9953	1.0000							
8	.0199	.5747	.7995	.9849	.9975	1.0000						
9	.0105	.4835	.7329	.9692	.9914	1.0000						
10	.0055	.4010	.6660	.9490	.9815	.9997	1.0000					
11	.0029	.3283	.6004	.9248	.9682	.9986	.9998	1.0000				
12	.0015	.2658	.5376	.8971	.9519	.9964	.9992	1.0000				
13	.0008	.2131	.4783	.8666	.9330	.9929	.9979	1.0000				
14	.0004	.1692	.4232	.8338	.9117	.9880	.9957	.9999	1.0000			
15	.0002	.1333	.3725	.7993	.8886	.9818	.9926	.9996	.9999	1.0000		
16	.0001	.1043	.3262	.7635	.8637	.9742	.9884	.9991	.9998	1.0000		
17	.0001	.0810	.2846	.7269	.8375	.9654	.9832	.9984	.9995	1.0000		
18	.0000	.0626	.2472	.6900	.8103	.9552	.9770	.9972	.9990	1.0000		
19	.0000	.0481	.2139	.6530	.7823	.9439	.9698	.9956	.9983	.9999	1.0000	
20	.0000	.0367	.1845	.6162	.7537	.9315	.9616	.9936	.9972	.9998	.9999	1.0000
21	.0000	.0279	.1585	.5800	.7248	.9181	.9526	.9911	.9959	.9996	.9999	1.0000
22	.0000	.0217	.1359	.5445	.6957	.9037	.9427	.9881	.9941	.9993	.9998	1.0000
23	.0000	.0160	.1161	.5100	.6666	.8885	.9320	.9845	.9920	.9990	.9996	1.0000
24	.0000	.0120	.0989	.4765	.6377	.8725	.9205	.9805	.9895	.9984	.9993	1.0000
25	.0000	.0090	.0841	.4442	.6091	.8558	.9084	.9759	.9866	.9977	.9990	.9999
26	.0000	.0067	.0713	.4133	.5809	.8386	.8957	.9709	.9833	.9969	.9985	.9998
27	.0000	.0050	.0603	.3837	.5531	.8208	.8824	.9654	.9796	.9958	.9979	.9998
28	.0000	.0037	.0509	.3555	.5260	.8025	.8686	.9593	.9754	.9946	.9972	.9996
29	.0000	.0027	.0429	.3288	.4996	.7839	.8543	.9529	.9709	.9931	.9963	.9994
30	.0000	.0020	.0361	.3035	.4738	.7649	.8396	.9459	.9660	.9915	.9953	.9992

n	1	2	3	4	5	6	7	8	9	10	11	12
31	.0000	.0015	.0303	.2797	.4489	.7457	.8246	.9386	.9607	.9896	.9941	.9989
32	.0000	.0011	.0254	.2573	.4247	.7264	.8092	.9308	.9550	.9875	.9927	.9985
33	.0000	.0008	.0212	.2363	.4014	.7069	.7936	.9226	.9490	.9851	.9911	.9981
34	.0000	.0006	.0177	.2167	.3790	.6873	.7777	.9141	.9427	.9825	.9894	.9976
35	.0000	.0004	.0148	.1985	.3574	.6677	.7617	.9052	.9360	.9797	.9874	.9970
36	.0000	.0003	.0123	.1815	.3366	.6481	.7455	.8959	.9290	.9766	.9852	.9963
37	.0000	.0002	.0102	.1657	.3168	.6286	.7291	.8864	.9217	.9733	.9829	.9955
38	.0000	.0002	.0085	.1511	.2979	.6092	.7128	.8765	.9141	.9698	.9804	.9946
39	.0000	.0001	.0071	.1376	.2800	.5899	.6963	.8664	.9062	.9661	.9777	.9935
40	.0000	.0001	.0059	.1251	.2625	.5708	.6799	.8560	.8981	.9621	.9748	.9924
41	.0000	.0001	.0048	.1137	.2462	.5520	.6634	.8454	.8898	.9579	.9717	.9912
42		.0000	.0040	.1031	.2306	.5334	.6471	.8345	.8817	.9535	.9684	.9898
43		.0000	.0033	.0935	.2158	.5150	.6307	.8235	.8724	.9489	.9649	.9883
44		.0000	.0027	.0846	.2018	.4970	.6145	.8122	.8633	.9440	.9612	.9867
45		.0000	.0023	.0765	.1886	.4793	.5984	.8008	.8541	.9390	.9574	.9850
46		.0000	.0019	.0691	.1761	.4618	.5824	.7893	.8448	.9339	.9534	.9832
47		.0000	.0015	.0624	.1643	.4448	.5665	.7776	.8352	.9284	.9492	.9812
48		.0000	.0013	.0562	.1532	.4281	.5509	.7658	.8255	.9229	.9448	.9791
49		.0000	.0010	.0506	.1427	.4118	.5354	.7539	.8157	.9170	.9403	.9769
50		.0000	.0008	.0456	.1329	.3958	.5201	.7419	.8057	.9111	.9356	.9745
51		.0000	.0007	.0410	.1236	.3803	.5050	.7299	.7957	.9050	.9308	.9721
52		.0000	.0006	.0368	.1150	.3652	.4901	.7177	.7855	.8987	.9258	.9695
53		.0000	.0005	.0330	.1068	.3504	.4755	.7056	.7752	.8923	.9207	.9668
54		.0000	.0004	.0296	.0992	.3361	.4611	.6934	.7649	.8858	.9154	.9640
55		.0000	.0003	.0265	.0921	.3222	.4470	.6812	.7545	.8791	.9100	.9611
56		.0000	.0003	.0237	.0854	.3087	.4331	.6689	.7440	.8722	.9045	.9580
57		.0000	.0002	.0212	.0791	.2956	.4195	.6567	.7335	.8653	.8988	.9548
58		.0000	.0002	.0190	.0733	.2829	.4061	.6445	.7229	.8582	.8930	.9516
59		.0000	.0001	.0170	.0679	.2707	.3930	.6324	.7123	.8510	.8871	.9482
60		.0000	.0001	.0151	.0628	.2588	.3802	.6202	.7017	.8437	.8811	.9447
61		.0000	.0001	.0135	.0581	.2473	.3677	.6082	.6911	.8363	.8750	.9411
62		.0000	.0001	.0120	.0537	.2363	.3555	.5961	.6805	.8288	.8687	.9374
63		.0000	.0001	.0107	.0496	.2256	.3435	.5842	.6698	.8212	.8624	.9336

n	3	4	5	6	7	8	9	10	11	12
64	.0000	.0095	.0458	.2153	.3319	.5722	.6592	.8135	.8560	.9297
65	.0000	.0085	.0423	.2054	.3205	.5604	.6486	.8057	.8495	.9257
66	.0000	.0076	.0390	.1958	.3094	.5487	.6380	.7979	.8429	.9216
67	.0000	.0067	.0360	.1867	.2986	.5371	.6275	.7900	.8362	.9174
68	.0000	.0060	.0332	.1778	.2880	.5256	.6169	.7820	.8295	.9131
69	.0000	.0053	.0306	.1694	.2778	.5141	.6065	.7740	.8227	.9087
70	.0000	.0047	.0281	.1612	.2678	.5028	.5960	.7660	.8158	.9042
71	.0000	.0042	.0259	.1534	.2581	.4916	.5857	.7577	.8088	.8997
72	.0000	.0037	.0238	.1459	.2487	.4806	.5753	.7495	.8018	.8950
73	.0000	.0033	.0219	.1388	.2396	.4696	.5651	.7413	.7948	.8903
74	.0000	.0029	.0202	.1319	.2307	.4588	.5549	.7331	.7877	.8855
75	.0000	.0026	.0185	.1253	.2221	.4481	.5448	.7248	.7805	.8807
76	.0000	.0023	.0170	.1190	.2137	.4376	.5348	.7165	.7733	.8757
77	.0000	.0020	.0156	.1130	.2056	.4272	.5248	.7081	.7660	.8707
78	.0000	.0018	.0143	.1072	.1978	.4170	.5149	.6998	.7588	.8657
79	.0000	.0016	.0132	.1017	.1902	.4069	.5051	.6914	.7514	.8605
80	.0000	.0014	.0121	.0965	.1828	.3969	.4955	.6831	.7441	.8553
81	.0000	.0012	.0111	.0915	.1757	.3871	.4859	.6747	.7367	.8501
83	.0000	.0011	.0101	.0867	.1688	.3775	.4764	.6663	.7293	.8448
83	.0000	.0009	.0093	.0814	.1621	.3680	.4670	.6579	.7219	.8394
84	.0000	.0008	.0085	.0778	.1557	.3587	.4577	.6496	.7145	.8340
85	.0000	.0007	.0078	.0737	.1495	.3495	.4485	.6412	.7070	.8285
86	.0000	.0006	.0071	.0697	.1435	.3405	.4394	.6329	.6996	.8229
87	.0000	.0006	.0065	.0660	.1377	.3317	.4304	.6245	.6921	.8174
88	.0000	.0005	.0060	.0624	.1321	.3230	.4215	.6162	.6847	.8117
89	.0000	.0004	.0055	.0590	.1267	.3144	.4128	.6079	.6772	.8061
90	.0000	.0004	.0050	.0558	.1215	.3061	.4041	.5997	.6697	.8004
91	.0000	.0003	.0046	.0527	.1164	.2979	.3956	.5914	.6623	.7946
92	.0000	.0003	.0042	.0498	.1116	.2898	.3872	.5832	.6548	.7888
93	.0000	.0003	.0038	.0471	.1069	.2820	.3789	.5751	.6473	.7830
94	.0000	.0002	.0035	.0444	.1024	.2742	.3707	.5669	.6399	.7772
95	.0000	.0002	.0032	.0420	.0981	.2667	.3627	.5588	.6325	.7713
96	.0000	.0002	.0029	.0396	.0940	.2593	.3547	.5508	.6251	.7654
97	.0000	.0002	.0026	.0374	.0900	.2520	.3469	.5428	.6176	.7595
98	.0000	.0001	.0024	.0353	.0861	.2449	.3392	.5348	.6103	.7535
99	.0000	.0001	.0022	.0332	.0824	.2380	.3316	.5269	.6029	.7475
100	.0000	.0001	.0020	.0313	.0788	.2312	.3242	.5190	.5956	.7415

$$Pr\{d_p^{(2)}(\mu_n,\nu_n) \le z/2n\}$$

n	z=1	2	3	4	5	6	7	8	9	10	11
1	1.0000										
2	1.0000										
3	.9000	1.0000									
4	.7429	1.0000									
5	.5714	.9841	1.0000								
6	.4156	.9546	1.0000								
7	.2890	.9103	.9977	1.0000							
8	.1939	.8546	.9918	1.0000							
9	.1264	.7899	.9819	.9997	1.0000						
10	.0804	.7194	.9678	.9986	1.0000						
11	.0501	.6462	.9493	.9965	1.0000						
12	.0307	.5728	.9266	.9930	.9998	1.0000					
13	.0185	.5016	.9000	.9881	.9994	1.0000					
14	.0110	.4343	.8698	.9816	.9986	1.0000	1.0000				
15	.0065	.3721	.8366	.9733	.9973	.9999	1.0000				
16	.0038	.3157	.8007	.9633	.9955	.9997	1.0000				
17	.0022	.2654	.7627	.9516	.9931	.9994	1.0000				
18	.0013	.2212	.7232	.9381	.9900	.9990	1.0000				
19	.0007	.1830	.6826	.9230	.9862	.9983	.9999	1.0000			
20	.0004	.1502	.6414	.9063	.9816	.9974	.9998	1.0000			
21	.0002	.1225	.6002	.8881	.9761	.9962	.9996	1.0000			
22	.0001	.0993	.5593	.8684	.9699	.9947	.9994	.9999	1.0000		
23	.0001	.0800	.5191	.8475	.9628	.9929	.9990	.9999	1.0000		
24	.0000	.0641	.4799	.8254	.9549	.9907	.9986	.9998	1.0000		
25	.0000	.0511	.4420	.8022	.9461	.9881	.9980	.9997	1.0000		
26	.0000	.0405	.4057	.7782	.9365	.9851	.9972	.9996	.9999	1.0000	
27	.0000	.0320	.3711	.7533	.9261	.9817	.9963	.9994	.9999	1.0000	
28	.0000	.0251	.3383	.7278	.9149	.9778	.9952	.9992	.9999	1.0000	
29	.0000	.0197	.3074	.7019	.9029	.9735	.9940	.9989	.9998	1.0000	

n	1	2	3	4	5	6	7	8	9	10	11
30	.0000	.0153	.2784	.6756	.8902	.9688	.9925	.9986	.9998	1.0000	
31	.0000	.0119	.2514	.6490	.8768	.9636	.9908	.9981	.9997	1.0000	
32	.0000	.0092	.2264	.6224	.8627	.9579	.9889	.9976	.9996	.9999	1.0000
33	.0000	.0071	.2034	.5958	.8480	.9518	.9868	.9969	.9994	.9999	1.0000
34	.0000	.0055	.1822	.5693	.8328	.9452	.9844	.9962	.9992	.9999	1.0000
35	.0000	.0042	.1627	.5431	.8170	.9382	.9818	.9954	.9990	.9998	1.0000
36	.0000	.0032	.1450	.5172	.8006	.9307	.9789	.9944	.9987	.9998	1.0000
37	.0000	.0024	.1289	.4917	.7838	.9228	.9757	.9933	.9984	.9997	.9999
38	.0000	.0019	.1144	.4667	.7666	.9144	.9723	.9921	.9980	.9996	.9999
39	.0000	.0014	.1012	.4423	.7491	.9057	.9686	.9908	.9976	.9995	.9999
40	.0000	.0011	.0894	.4185	.7312	.8965	.9647	.9892	.9971	.9994	.9999
41	.0000	.0008	.0788	.3954	.7131	.8869	.9605	.9876	.9966	.9992	.9998
42	.0000	.0006	.0693	.3730	.6948	.8770	.9560	.9859	.9960	.9990	.9998
43	.0000	.0005	.0608	.3513	.6763	.8666	.9512	.9839	.9953	.9988	.9997
44	.0000	.0003	.0533	.3305	.6770	.8559	.9462	.9818	.9945	.9985	.9997
45	.0000	.0003	.0466	.3104	.6390	.8449	.9409	.9796	.9936	.9983	.9996
46	.0000	.0002	.0407	.2912	.6203	.8336	.9353	.9772	.9927	.9979	.9995
47	.0000	.0001	.0355	.2728	.6016	.8219	.9295	.9746	.9917	.9976	.9994
48	.0000	.0001	.0308	.2553	.5829	.8100	.9234	.9719	.9906	.9972	.9992
49	.0000	.0001	.0268	.2385	.5643	.7978	.9170	.9689	.9894	.9967	.9991
50	.0000	.0001	.0232	.2226	.5459	.7853	.9104	.9659	.9881	.9962	.9990
51		.0000	.0201	.2075	.5276	.7726	.9036	.9626	.9867	.9957	.9987
52		.0000	.0174	.1932	.5094	.7598	.8965	.9592	.9852	.9951	.9985
53		.0000	.0150	.1797	.4915	.7467	.8891	.9556	.9836	.9945	.9983
54		.0000	.0130	.1669	.4739	.7334	.8815	.9518	.9819	.9938	.9980
55		.0000	.0112	.1549	.4565	.7200	.8737	.9479	.9801	.9930	.9977
56		.0000	.0096	.1436	.4394	.7065	.8657	.9438	.9782	.9922	.9974
57		.0000	.0082	.1329	.4226	.6928	.8575	.9395	.9761	.9913	.9971
58		.0000	.0071	.1230	.4061	.6791	.8490	.9350	.9740	.9904	.9967
59		.0000	.0061	.1136	.3900	.6653	.8404	.9304	.9718	.9894	.9963
60		.0000	.0052	.1049	.3742	.6514	.8316	.9256	.9694	.9883	.9959

n	1	2	3	4	5	6	7	8	9	10	11
61		.0000	.0044	.0967	.3588	.6375	.8226	.9207	.9669	.9872	.9954
62		.0000	.0038	.0891	.3438	.6238	.8134	.9156	.9643	.9860	.9949
63		.0000	.0032	.0820	.3292	.6097	.8040	.9103	.9617	.9847	.9943
64		.0000	.0028	.0755	.3150	.5958	.7945	.9048	.9588	.9834	.9937
65		.0000	.0023	.0693	.3012	.5819	.7849	.8992	.9559	.9819	.9931
66		.0000	.0020	.0636	.2878	.5681	.7751	.8935	.9529	.9805	.9924
67		.0000	.0017	.0584	.2748	.5543	.7652	.8876	.9497	.9789	.9917
68		.0000	.0014	.0535	.2622	.5406	.7551	.8816	.9465	.9773	.9909
69		.0000	.0012	.0490	.2501	.5270	.7450	.8754	.9431	.9756	.9901
70		.0000	.0010	.0448	.2383	.5134	.7347	.8690	.9396	.9738	.9892
71		.0000	.0009	.0410	.2270	.5000	.7244	.8626	.9360	.9719	.9884
72		.0000	.0007	.0374	.2161	.4868	.7140	.8560	.9323	.9700	.9874
73		.0000	.0006	.0342	.2055	.4736	.7035	.8492	.9284	.9680	.9864
74		.0000	.0005	.0312	.1954	.4606	.6929	.8424	.9245	.9659	.9854
75		.0000	.0004	.0284	.1856	.4478	.6823	.8354	.9204	.9637	.9843
76		.0000	.0004	.0289	.1763	.4351	.6716	.8283	.9163	.9615	.9832
77		.0000	.0003	.0236	.1673	.4226	.6609	.8211	.9120	.9591	.9820
78		.0000	.0003	.0214	.1587	.4103	.6501	.8137	.9076	.9567	.9807
79		.0000	.0002	.0195	.1502	.3982	.6394	.8063	.9032	.9542	.9794
80		.0000	.0002	.0177	.1425	.3863	.6286	.7988	.8986	.9517	.9781
81		.0000	.0002	.0161	.1350	.3745	.6178	.7912	.8939	.9490	.9767
82		.0000	.0001	.0146	.1277	.3630	.6070	.7834	.8891	.9463	.9752
83		.0000	.0001	.0132	.1208	.3517	.5962	.7756	.8842	.9435	.9740
84		.0000	.0001	.0120	.1142	.3406	.5854	.7677	.8792	.9406	.9721
85		.0000	.0001	.0108	.1079	.3297	.5747	.7598	.8741	.9377	.9705
86		.0000	.0001	.0098	.1019	.3191	.5640	.7517	.8690	.9346	.9688
87		.0000	.0001	.0089	.0962	.3086	.5533	.7436	.8637	.9315	.9671
88			.0000	.0080	.0908	.2984	.5427	.7354	.8583	.9283	.9653
89			.0000	.0072	.0856	.2884	.5321	.7272	.8529	.9250	.9635

n	1	2	3	4	5	6	7	8	9	10	11
90			.0000	.0065	.0807	.2787	.5216	.7189	.8474	.9217	.9616
91			.0000	.0059	.0761	.2691	.5111	.7105	.8417	.9183	.9596
92			.0000	.0053	.0716	.2599	.5008	.7021	.8360	.9148	.9576
93			.0000	.0048	.0674	.2508	.4904	.6937	.8303	.9112	.9556
94			.0000	.0043	.0634	.2420	.4802	.6852	.8244	.9075	.9534
95			.0000	.0039	.0597	.2333	.4700	.6767	.8185	.9038	.9512
96			.0000	.0035	.0561	.2250	.4600	.6682	.8125	.9000	.9490
97			.0000	.0032	.0527	.2168	.4500	.6596	.8064	.8962	.9467
98			.0000	.0028	.0495	.2089	.4402	.6510	.8002	.8922	.9444
99			.0000	.0025	.0465	.2012	.4304	.6424	.7940	.8882	.9420
100			.0000	.0023	.0436	.1937	.4207	.6338	.7878	.8842	.9395

A.3 Bounds on $\Pr\{f(W^\circ) \le c\}$.

Proposition 4.1.6 gives as upper and lower bounds on $\Pr\{f(W^\circ) \le c\}$ the values $\Pr\{T_1(W^\circ) \le 2c,\ T_2(W^\circ) \le 3c\}$ and $\Pr\{T_2(W^\circ) \le 2c\}$. Equivalently, we may write $\Pr\{(m^\circ, M^\circ) \in B\}$ and $\Pr\{(m^\circ, M^\circ) \in A\}$ where

$$A = \{(x,y):\ x \le 0,\ y \ge 0 \text{ and } y-x \le 2c\} \quad \text{and}$$

$$B = \{(x,y):\ -2c \le x \le 0,\ 0 \le y \le 2c \text{ and } y-x \le 3c\},$$

where m° and M° are as in 1.4.3. 1.4.3(c) permits the evaluation of $\Pr\{(m^\circ, M^\circ) \in \{(x,y) = 0 \ge x \ge a,\ 0 \le y \le b\}\}$. (The conditions $0 \ge x$ and $0 \le y$ are unnecessary since $m^\circ \le 0$ and $M^\circ \ge 0$.)

A and B can be approximated arbitrarily closely by sets of the form $\{(x,y) = x \ge a,\ y \le b\}$. For example

$$A \subseteq \Big[\bigcup_{i=1}^{2n} \{(x,y):\ -\frac{ic}{n} \le x \le \frac{-(i-1)c}{n}\ ; \quad -\infty < y \le 2c - \frac{(i-1)}{n}c\}\Big] - \{-2c \le x \le 0,\ -\infty < y \le 0\}\ .$$

$$A \supseteq \Big[\bigcup_{i=1}^{2n-1} \{(x,y):\ -\frac{ic}{n} \le x \le -\frac{(i-1)}{n}c;\ \ -\infty < y < 2c - \frac{(i+1)}{n}c\}\Big] - \{-2c \le x \le 0,\ -\infty < y \le 0\}\ .$$

Each set of the form $\{-\frac{i}{n}c \le x \le -\frac{(i-1)}{n}c;\ -\infty < y < 2c - \frac{(i-1)}{n}c\}$ may be expressed as

$$\{-\frac{i}{n}c \le x;\ -\infty < y < 2c - \frac{(i-1)}{n}c\} - \{-\frac{(i-1)}{n}c \le x \le 0;\ -\infty < y < 2c - \frac{(i-1)c}{n}\},$$

so that 1.4.3(c) can be used to compute its probability. The set B can be approximated in a similar manner.

We give the upper and lower bounds computed in this way for selected values of c. The intervals between values of c are chosen to be smaller where the distribution functions are increasing most rapidly. Compare the upper bound for $\Pr\{f(W^{\circ}) \leq c\}$ with table A1 for $n = 40$, 60 and 80.

c	$\Pr\{T_2(W^\circ) \le 2c\}$ (lower bound)	$\Pr\{T_1(W^\circ) \le 2c,\ T_2(W^\circ) \le 3c\}$ (upper bound)
.225		.001
.25		.006
.275		.022
.3		.054
.325	.000	.107
.35	.002	.178
.375	.009	.263
.4	.021	.356
.425	.043	.447
.45	.075	.536
.5	.178	.690
.55	.314	.802
.6	.465	.879
.65	.609	.929
.7	.728	.959
.75	.821	.977
.8	.889	.988
.85	.934	.994
.9	.963	.997
.95	.979	.999
1.00	.989	.999

BIBLIOGRAPHY

[1] Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist. 23*, 193-212.

[2] Berge, C. (1966). *Theory of Graphs and its Applications.* Methuen Co. Litd., London.

[3] Bhattacharya, R.N. (1971). Rates of weak convergence and asymptotic expansions for classical central limit theorems. *Ann. Math. Statist. 42*, 241-259.

[4] Bickel, P.J. (1969). A distribution free version of the Smirnov two sample test in the p-variate case. *Ann. Math. Statist. 40*, 1-23.

[5] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[6] Birnbaum, Z.W. and Tingey, F.H. (1951). One-sided confidence contours for probability distribution functions. *Ann. Math. Statist. 22*, 592-596.

[7] Birnbaum, Z.W. (1952). Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size. *J. Amer. Stat. Assn. 47*, 425-441.

[8] Darling, D.A. (1957). The Kolmogorov-Smirnov, Cramér-von Mises tests. *Ann. Math. Statist.,28* 823-838.

[9] Donsker, M. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist. 23*, 277-281.

[10] Doob, J.L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorem. *Ann. Math. Statist. 20*, 393-403.

[11] Dudley, R.M. (1968). Distances of probability measures and random variables. *Ann. Math. Statist. 39*, 1563-1572.

[12] Dudley, R.M. (1969). The speed of mean Glivenko-Cantelli convergence. *Ann. Math. Statist. 40*, 40-50.

[13] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Vol. I.* Wiley, New York.

[14] Gnedenko, B.V. and Kolmogorov, A.N. (1954). *Limit Distributions for Sums of Independent Random Variables*. (transl. by K.L. Chung). Addison-Wesley, Reading, Mass.

[15] Hampel, F.R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist. 42*, 1887-1896.

[16] Hopcroft, J.E. and Karp, R.M. (1973). An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM J. Comp. 2*, 225-231.

[17] Kiefer, J. (1959). k-sample analogues of the Kolmogorov-Smirnov and Cramér-von Mises tests. *Ann. Math. Statist. 30*, 420-447.

[18] Kuhn, H.W. (1955). The Hungarian method for the Assignment Problem. *Naval Res. Logist. Quart. 2*, 83-97.

[19] Lehmann, E.L. (1953). The power of rank tests. *Ann. Math. Statist. 24*, 23-43.

[20] Massey, F.J., Jr. (1951). The distribution of the maximum deviation between two sample cumulative step functions. *Ann. Math. Statist., 22*, 125-128.

[21] Massey, F.J., Jr. (1952). Distribution table for the deviation between sample cumulatives. *Ann. Math. Statist. 23*, 435-441.

[22] Massey, F.J., Jr. (1952). Correction to "A note on the power of a non-parametric test." *Ann. Math. Statist. 23*, 637-638.

[23] Ore, O. (1955). Graphs and matching theorems. *Duke Math. J. 22*, 625-639.

[24] Prokhorov, Hu. (1956). Convergence of random processes and limit theorems. *Theor. Probability Appl. 1*, 157-214.

[25] Rosenblatt, J. (1962). Some modified Kolmogorov-Smirnov tests of approximate hypotheses and their properties. *Ann. Math. Statist. 33*, 513-524.

[26] Rosenblatt, J. (1963). Tests and confidence intervals based on the metric d_2. *Ann. Math. Statist. 34*, 618-623.

[27] Rosenkrantz, W.A. (1969). A rate of convergence for the von Mises statistic. *Trans. Amer. Math. Soc. 139*, 329-337.

[28] Skorokhod, A.V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl. 1*, 261-290.

[29] Smirnov, N.V. (1948). Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Statist. 19*, 279-281.

[30] Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist. 36*, 423-439.

[31] Varadarajan, V.S. (1958). On the convergence of sample probability distributions. *Sankhyā 19*, 23-26.

[32] Zolotarev, V.M. (1971). Estimates of the difference between distributions in the Lèvy metric. *Proc. Steklov Inst. Math. 112*, 232-240.