The Annals of Probability
1974, Vol. 2, No. 3, 501-508

GENERALIZED DISTRIBUTION FUNCTIONS: THE LINEARLY ORDERED CASE WITH APPLICATIONS TO NONPARAMETRIC STATISTICS[1]

BY GORDON SIMONS

University of North Carolina

We develop a theory for distribution functions which are defined on an abstract linearly ordered space. This extends and gives additional insight into the work of J. H. B. Kemperman who was concerned with a special class of linearly ordered spaces. Attention is given to nonparametric applications. The utility of Kemperman's work on tolerance regions and of related applications to goodness of fit tests, appearing in this paper, is enhanced. In particular, it is shown that a number of standard nonparametric procedures can be extended to cover almost any kind of random sample (e.g., multivariate and most time series data) occurring in practice.

1. Introduction and summary. The spaces studied by modern probabilists have become very general. Frequently some structure is imposed on a space, but typically not an ordering. Since distribution functions can only be defined on partially ordered spaces (classically, on finite dimensional Euclidean spaces), they have become the "poor cousins" of probability measures. Nevertheless, distribution functions serve several important roles: (a) For the probabilist, they *concisely* conserve the information represented by a probability measure on a finite dimensional Euclidean space. (b) For the statistician, among other things, they enter into the definition of certain goodness of fit test statistics such as the Kolmogorov statistic D_n.

In this paper we only consider distribution functions (df) for linearly ordered spaces. We present what we believe to be a reasonably complete theory for abstract linearly ordered spaces. An effort is made to find parallels with results for the real line. We find necessary and sufficient conditions for a df to uniquely determine the probability measure giving rise to it.

The theory specializes for a class of spaces introduced by J. H. B. Kemperman (1956). He was studying the subject of nonparametric tolerance regions. We consider alternative applications—nonparametric goodness of fit tests—in a slightly more general setting than his. On the practical side, Theorem 3 below permits one to apply a wide range of nonparametric statistical procedures to many new situations. What is required is a random sample with the observations assuming values in a separable metric space. This includes multivariate data and many types of times series data.

Received September 1, 1972; revised October 1, 1973.

[1] This research was supported in part by the U.S. Air Force Under Contract No. AFOSR-68-1415.

AMS 1970 *subject classifications.* Primary 60E05, 62E10, 62E15, 62G10; Secondary 06A05.

Key words and phrases. Distribution function, linear ordering, nonparametric statistics.

We hope, in subsequent studies, to follow up on these applications and to consider df's on partially ordered spaces. The current study was motivated by the work of Flavio Rodrigues (1972) who used generalized distribution functions while studying relationships between weak convergence and probability convergence.

2. Distribution functions on linearly ordered spaces. Let $\mathscr{X}$ be a linearly ordered space. We introduce the following notions: $A \subseteqq \mathscr{X}$ is an *initial* (*terminal*) if $x < y \in A$ ($x > y \in A$) entails $x \in A$. An *interval* is the intersection of an initial and a terminal. The smallest σ-field $\mathscr{B}$ containing the intervals (equivalently, the initials) of $\mathscr{X}$ is the *order σ-field* and its members are *order sets*. $\bar{x}$ ($\vec{x}$) denotes the initial $\{y : y \leqq x\}$ (terminal $\{y : y \geqq x\}$) and is called a *closed* initial (terminal). Let P be a probability measure (p.m.) on $(\mathscr{X}, \mathscr{B})$. The *distribution function* (df) of P is the function F defined by $F(x) = P(\bar{x})$, $x \in \mathscr{X}$.

For the real line, $\mathscr{B}$ is the Borel σ-field and F (uniquely) determines the values of P on $\mathscr{B}$ for every P. The following theorem tells us when this important property holds for a given $\mathscr{X}$ and df F.

THEOREM 1. (i) *F determines P on $\mathscr{B}$ if it determines $P(A)$ for each initial A.*

(ii) *F determines $P(A)$ for an initial A if and only if either* (a) *A is expressible in the form $\bigcup_{n=1}^{\infty} \bar{x}_n$, $\bigcap_{n=1}^{\infty} \bar{x}$, $\mathscr{X}$ or $\emptyset$ (the empty set), or* (b) $\sup_{x \in A} F(x) = \inf_{y \notin A} F(y)$.

(iii) *It is possible for F not to determine P on $\mathscr{B}$.*

PROOF. To see (i), observe that F will determine P on the intervals of $\mathscr{X}$, a semi-ring. The "if" part of (ii) is straightforward. Therefore, suppose (a) and (b) do not hold and define a set function Q on the initials satisfying $Q(B) = P(B)$ if $B \neq A$ and $Q(A) \in [\sup_{x \in A} F(x), \inf_{y \notin A} F(y)]$ with $Q(A) \neq P(A)$. Clearly, F will be the df of any extension of Q which is a p.m. on $(\mathscr{X}, \mathscr{B})$. It remains to show such an extension exists. The obvious extension of Q to the intervals is unique and nonnegative. Suppose $I = \sum_{n=1}^{\infty} I_n$, where I and each I_n is an interval. The needed equality $Q(I) = \sum_{m=1}^{\infty} Q(I_n)$ is inherited from P except possibly when I is the difference of two initials $B_2 - B_1$ one of which is A. Suppose $B_2 = A$. Then there exists a largest interval I_{n_0}, say. Otherwise, by choosing an x_n in each I_n, one would obtain the contradiction $A = \bigcup_{n=1}^{\infty} \bar{x}_n$. Then the interval $I - I_{n_0} = \sum_{n=1, n \neq n_0}^{\infty} I_n$, and the equation $Q(I) = \sum_{m=1}^{\infty} Q(I_n)$ is, in fact, inherited from P. The case $B_1 = A$ is handled similarly. To show (iii), let $\mathscr{X} = \{\text{ordinals} \leqq \Omega\}$, where Ω is the first uncountable ordinal, and $P(\{\Omega\}) = 1$. The initial $A = \mathscr{X} - \{\Omega\}$ fails to satisfy condition (a) or (b). (A second p.m. with the same df is given in the proof of Proposition 1 below.) $\square$

Theorem 1 is intentionally expressed in as usable a form as possible. The conditions do not depend on P. Some insight can be gained by noting that condition (a) characterizes the initials in the σ-field generated by the closed initials.

The behaviour of F and its relationship to P largely depends on the atomic structure of the probability space $(\mathscr{X}, \mathscr{B}, P)$. (For background, see Neveu

(1965), page 18.) There are two types of atoms possible—point atoms and interval atoms which are not also point atoms: An atom A will be called a *point atom* (*interval atom*) if there exists a point x (interval I) such that $P(A \bigtriangleup \{x\}) = 0$ ($P(A \bigtriangleup I) = 0$).

PROPOSITION 1. (i) *Every atom of* $(\mathscr{X}, \mathscr{B}, P)$ *is an interval atom.*
(ii) *Non-point atoms are possible.*

PROOF. Using Theorem D, page 56, of Halmos (1950), one can closely approximate an atom A by a finite union of disjoint intervals. Since A is an atom, this approximation (when good enough) holds for one of the intervals in the union. The interval needed for the proof of (i) is obtained as the intersection of a sequence of such approximations. To show (ii), let $\mathscr{X} = \{\text{ordinals} \leqq \Omega\}$ where Ω is the first uncountable ordinal. Further, let $P(A) = 0$ or 1 for each $A \in \mathscr{B}$ as A contains a countable or uncountable number of points. It is easily checked that P is a p.m. and $\mathscr{X}$ is a non-point atom. $\square$

According to Neveu (1965), page 18, $\mathscr{X}$ can be decomposed into a countable union of atoms and an atomless part. With this and Proposition 1, we can easily obtain:

PROPOSITION 2. *P can be decomposed uniquely as $\alpha_1 P_1 + \alpha_2 P_2 + \alpha_3 P_3$, with $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and each $\alpha_i \geqq 0$, where P_1 is a nonatomic* p.m. (*when* $\alpha_1 > 0$), *where P_2 is a completely point atomic* p.m. (*when* $\alpha_2 > 0$), *and where P_3 is a completely non-point interval atomic* p.m. (*when* $\alpha_3 > 0$).

We shall find that F behaves in much the same way as a classical df when $\alpha_3 = 0$. Throughout the remainder of the paper, we shall link properties of F with the values of the α_i's. We shall be particularly interested in finding verifiable conditions which guarantee that $\alpha_3 = 0$.

The analog of a continuous df is a *dense* df, a df whose range is dense in $[0, 1]$. If F is dense, $\alpha_3 = 0$:

PROPOSITION 3. *F is dense if and only if P is nonatomic* ($\alpha_1 = 1$).

PROOF. The "only if" part is obvious. Therefore, suppose P is nonatomic. Let C be the range of P on the initials of $\mathscr{X}$. It suffices to show $C = [0, 1]$. It is easy to check that C is a closed set. Now suppose $(P(A_1), P(A_2))$ is a gap in C for initials A_1 and A_2. Then the p.m. $Q(\cdot) = P(\cdot)/P(A_2 - A_1)$ on $(A_2 - A_1, \mathscr{B}(A_2 - A_1))$ is 0 or 1 for every interval in $A_2 - A_1$ and, hence, on $\mathscr{B}(A_2 - A_1)$. This untenable situation would make $A_2 - A_1$ an atom for $(\mathscr{X}, \mathscr{B}, P)$. $\square$

F is said to be *discrete* if it can be expressed as $F(x) = \sum_{x_n \leq x} p_n$, $x \in \mathscr{X}$, where each $p_n > 0$ and the x_n are distinct. It does not automatically follow that ($P = P_2$ and) $p_n = P(\{x_n\})$. (See, for instance, the example in the proof of Proposition 1.) An initial (terminal) is *open* if its complement is a closed terminal (initial). F is *right continuous at a point* x if, for each $\varepsilon > 0$, there exists an open initial A with $x \in A$ and $F(u) < F(x) + \varepsilon$ for each $u \in A$. F is *right continuous* if it is right continuous at each point x for which $\bar{x} \neq \mathscr{X}$. $\alpha_1 = 0$ does not imply F is discrete. Neither must F be right continuous:

PROPOSITION 4. *Suppose* $\alpha_3 = 0$. *Then*:

(i) *$F(x)$ can be (uniquely) decomposed as $\alpha_1 F_1(x) + \alpha_2 F_2(x)$, where $F_1(x) = P_1(\bar{x})$ is a dense* df *and $F_2(x) = P_2(\bar{x})$ is a discrete* df *of the form $\sum_{x_n \leq x} P_2(\{x_n\})$, where $\{x_n\}_{n \geq 1}$ denote the point atoms of P.*

(ii) *F determines P on $\mathscr{B}$.*

(iii) *F is right continuous.*

For the sake of brevity, we omit the proof of this proposition. The proof depends on Proposition 3. It is very tempting to conclude from (ii) that the condition $\alpha_3 = 0$ entails (a) or (b) of Theorem 1 for each initial A of $\mathscr{X}$. This is *not* logically necessary and is not true.

Let $(\mathscr{Y}, \mathscr{C}, Q)$ be a second probability space where $\mathscr{Y}$ is linearly ordered and $\mathscr{C}$ is its order σ-field. $\mathscr{X} \times \mathscr{Y}$ can be linearly ordered lexicographically. Its order σ-field $\mathscr{D}$ is a sub σ-field of the product σ-field. Thus one obtains a probability space $(\mathscr{X} \times \mathscr{Y}, \mathscr{D}, R)$ where R is the (unique) product measure restricted to $\mathscr{D}$. Since there are three p.m.'s under consideration, we shall express the α_1 for P as $\alpha_1(P)$, etc. We shall need parts of the next proposition in later sections.

PROPOSITION 5. (i) *$\alpha_1(P) = 1$ entails $\alpha_1(R) = 1$.*

(ii) *$\alpha_3(P) = \alpha_3(Q) = 0$ entails $\alpha_3(R) = 0$.*

(iii) *The point (x, y) is a point atom of R if and only if x is a point atom of P and y is a point atom of Q.*

(iv) *Thus $\alpha_3(P) = 0$ and $\alpha_1(Q) = 1$ together entail $\alpha_1(R) = 1$.*

PROOF. Let $F^0(x) = P(\bar{x} - \{x\})$, $x \in \mathscr{X}$, and G be the df of R. Then

$$F^0(x) \leqq G(x, y) \leqq F(x), \qquad x \in \mathscr{X}, y \in \mathscr{Y}. \tag{1}$$

If F is dense then $F^0 = F$ and G is dense. Thus (i) follows from Proposition 3. (ii)—(iv) are easily proven. □

3. Kappa spaces. J. H. B. Kemperman (1956) originated and worked with a linearly ordered space[2] which we now refer to as a kappa space (κ space, for short). It will be seen that any linearly ordered space with weaker structure can have non-point atoms ($\alpha_3 > 0$). In this work, he also assumed that there were no point atoms ($\alpha_2 = 0$). Consequently ($\alpha_1 = 1$ and) F was dense. We find in the next section that this implies the probability integral transformation $F(X)$ is uniformly distributed on [0, 1], which is what Kemperman wanted.

An initial (terminal) is *approachable* if it can be expressed as a countable union of closed initials (terminals). (The empty set $\varnothing$ is approachable. The index set of the union is empty.) A linearly ordered space $\mathscr{X}$ is a *kappa space* if each initial and terminal in $\mathscr{X}$ is approachable.

[2] Actually he was working with a slightly more general ordering by linearly ordering equivalence classes of points. All of our results are expressible in his greater generality.

THEOREM 2. *$\mathscr{X}$ is a κ space if and only if $\alpha_3 = 0$ for each* p.m. *P on $(\mathscr{X}, \mathscr{B})$.*

PROOF. Suppose $\mathscr{X}$ is a κ space and A is an atom relative to P. We claim $P(BA) = 0$ for the initial $B = \{x: P(\bar{x}A) = 0\}$. For if $B \neq \emptyset$, then $B = \bigcup_{n=1}^{\infty} \bar{x}_n$, where we can assume $x_1 \leqq x_2 \leqq \cdots$. Then $P(BA) = \lim_{n\to\infty} P(\bar{x}_n A) = 0$. Similarly, $P(CA) = 0$ where $C = \{x: P(\bar{x}A) = 0\}$. Since $P(A) > 0$, there exists an $x \notin B \cup C$. Then $P(\bar{x}A) = P(\bar{x}A) = P(A)$, and it follows that $P(A \triangle \{x\}) = 0$. That is, A is a point atom. Conversely, suppose $\mathscr{X}$ is not a κ space. We shall illustrate the (representative) case where a nonempty initial A is unapproachable. Let $\mathscr{G}_0$ ($\mathscr{G}_1$) be the collection of $B \in \mathscr{B}$ which exclude (include) a top piece of A, a piece of the form $\bar{x}A$ with $x \in A$, and let $\mathscr{G} = \mathscr{G}_0 \cup \mathscr{G}_1$. $\mathscr{G}_0$ is closed under countable unions. For suppose $B = \bigcup_{n=1}^{\infty} B_n$, where each $B_n \in \mathscr{G}_0$. For each n, there exists an $x_n \in A$ for which $\bar{x}_n AB_n = \emptyset$. Since A is unapproachable, $A - \bigcup_{n=1}^{\infty} \bar{x}_n$ contains a point x and $\bar{x}AB = \emptyset$. Thus $B \in \mathscr{G}_0$. It easily follows that: $\mathscr{G}$ is a σ-field, $\mathscr{G} = \mathscr{B}$, and the set function P which is 0 on $\mathscr{G}_0$ and 1 on $\mathscr{G}_1$ is a p.m. $P(A) = 1$, so A is an atom. The unapproachability of A insures that it is a non-point atom. □

The importance of Theorem 2 is that it permits one to immediately make useful conclusions from the general theory of Section 2. For instance, for a κ space,

(i) F determines P,
(ii) F is dense if and only if $P(\{x\}) = 0$ for each $x \in \mathscr{X}$, and
(iii) F is discrete if and only if $\sum_{\{m \geqq 1\}} P(\{x_m\}) = 1$ for some countable set of distinct points $\{x_m\}_{m \geqq 1}$.

The real line is a κ space.

Let P be a p.m. on $(S, \mathscr{S})$ where S is a separable metric space and $\mathscr{S}$ is its Borel σ-field. An important consequence of the next theorem is as follows: S can be linearly ordered in such a way that P determines a df F and, conversely, F determines P on $\mathscr{S}$.

THEOREM 3. *S can be linearly ordered as a κ space in such a way that the order σ-field $\mathscr{B} = \mathscr{S}$.*

PROOF. (a) (based upon a proof by Skorokhod (1956) page 281) Let $|A|$ denote the diameter of a set $A \in \mathscr{S}$. One can easily produce a sequence of progressively refining countable partitions $Z_m = \{A_{i_1,i_2,\cdots,i_m}\}$ $(m \geqq 1)$ of S for which: (i) each element of each partition belongs to $\mathscr{S}$, (ii) $\sup\{|A|: A \in Z_m\} \to 0$ as $m \to \infty$, (iii) $A_{i_1,\cdots,i_m} = \bigcup_{\{i_{m+1}\}} A_{i_1,\cdots,i_{m+1}}$ $(m \geqq 1)$, and (iv) (for convenience) each index i_m is a positive integer.

Each point $x \in S$ can be identified uniquely with an infinite sequence of positive integers $i_1, i_2, i_3, \cdots$ according to its membership in each set $A_{i_1,\cdots,i_m}$, $m \geqq 1$. In turn, S can be linearly ordered lexicographically based on these infinite sequences. To avoid confusion, we shall provide S with the alternative name $\mathscr{X}$ which we shall use when referring to this linear ordering. $\mathscr{B}$ is the order σ-field associated with $\mathscr{X}$.

(b) *Showing $\mathscr{X}$ is a κ space*: Briefly, any nonempty non-closed initial (terminal) A can be expressed as the countable union $\bigcup \{\bar{x}_B : B \in Z_m$ for some $m \geqq 1$, $B \subseteq A\}$ ($\bigcup \{\bar{x}_B : B \in Z_m$ for some $m \geqq 1$, $B \subseteq A\}$), where x_B denotes an arbitrary point in the set B.

(c) *Showing $\mathscr{B} \subseteq \mathscr{S}$*: Since $\mathscr{X}$ is a κ space, the closed initials generate $\mathscr{B}$ and it suffices to show an arbitrary closed initial $\bar{x} \in \mathscr{S}$. But $\bar{x} = \bigcap_{m=1}^{\infty} B_m$, where B_m is the countable union of all sets in Z_m (and consequently members of $\mathscr{S}$) which precede or contain the point x.

(d) *Showing $\mathscr{S} \subseteq \mathscr{B}$*: Observe that each element C of each partition Z_m is an interval of $\mathscr{X}$, and hence belongs to $\mathscr{B}$. Thus it suffices to observe that each open ball $B \in \mathscr{S}$ is expressible as the countable union $\bigcup \{C: C \in Z_m$ for some $m \geqq 1$, $C \subseteq B\}$. □

Suppose, for example, that S is the open unit square. While the proof of Theorem 3 shows that there are many linear orderings of S which satisfy the requirements stated in the theorem, there is one which is particularly easy to describe. We can linearize S by mapping the point $(x, y) \in S$ into the decimal $\cdot x_1 y_1 x_2 y_2 \cdots$, where $x = \cdot x_1 x_2 \cdots$ and $y = \cdot y_1 y_2 \cdots$. Of course, we can linearize S lexicographically. But this would be unsatisfactory here since the resultant order σ-field $\mathscr{B}$ would be smaller than $\mathscr{S}$.

The reader can find some additional material on κ spaces in Simons (1972).

4. Probability integral transformations. Let X be a measurable mapping from some probability space to $(\mathscr{X}, \mathscr{B})$, which we will call a *random variable* (rv) *in* $\mathscr{X}$, and let F be its df (i.e., the df of the induced p.m. P on $(\mathscr{X}, \mathscr{B})$). $F(X)$ is a real random variable (rrv) since F is a measurable mapping from $(\mathscr{X}, \mathscr{B})$ to the real line. It seems appropriate, for historical reasons, to refer to this rrv as a *probability integral transformation*. Kemperman (1956) has studied this rrv in the context of a κ space and used it in establishing nonparametric tolerance regions. We shall study it in a somewhat broader context and, in the next section, discuss two nonparametric goodness of fit tests based upon it.

PROPOSITION 6. (i) *$F(X)$ is uniformly distributed on* $[0, 1]$ *if and only if F is dense.*

(ii) *If $\alpha_3 = 0$, $F(X)$ is stochastically no smaller than a uniformly distributed* rrv *on* $[0, 1]$.

PROOF. If F is dense, then for any $u \in \mathscr{X}$ for which $F(u) \in [0, 1)$ and any $\varepsilon > 0$, there exists a $v \in \mathscr{X}$ such that $F(u) < F(v) \leqq F(u) + \varepsilon$. Thus $F(u) = P(x : x \leqq u) \leqq P(x : F(x) \leqq F(u)) \leqq P(x : F(x) < F(v)) \leqq P(x : x \leqq v) = F(v) \leqq F(u) + \varepsilon$. Since $\varepsilon > 0$ is arbitrary, it follows that $P(x : F(x) \leqq t) = t$ for every t in the (dense) range of F and, hence, that $F(X)$ is uniformly distributed on $[0, 1]$. The converse is immediate. To show (ii), we refer to Proposition 5 and the notation used in connection with it. By choosing, $(\mathscr{Y}, \mathscr{C}, Q)$ so that

$\alpha_1(Q) = 1$, we see from part (iv) of Proposition 5, that $\alpha_1(R) = 1$; from Proposition 3 that G is dense; and, in turn, from part (i) of this proposition, that the corresponding probability integral transformation $G(X, Y)$ is uniformly distributed on [0, 1]. The conclusion is immediate from (1). □

REMARK. Likewise, from (1), we see that $F^0(X)$ is stochastically no larger than a uniformly distributed rrv on [0, 1].

5. Some applications to nonparametric statistics. Let $X_1, \cdots, X_n$ be independent random variables in $\mathscr{X}$ with each possessing a common unknown p.m. P and related df F. Unless F_0 determines P_0, we must distinguish between the hypothesis that $P = P_0$ and the hypothesis that $F = F_0$, the df of P_0. The latter type hypothesis requires test statistics defined in terms of F_0, not P_0, when the distinction must be made. We shall find that a generalized Kolmogorov statistic is easily defined in terms of F_0 but, unfortunately, a generalized Smirnov statistic requires P_0. (The issue is not very important here since F_0 determines P_0 for parts (ii) and (iii) of the next two theorems.)

Let $F_n(x) = n^{-1} \sum_{k=1}^{n} I_{[X_k \leq x]}$, $x \in \mathscr{X}$, denote the empirical df. The generalized Kolmogorov statistic is $D_n = \sup_{x \in \mathscr{X}} |F_n(x) - F_0(x)|$. We define the generalized Smirnov statistic as $W_n = \int_{\mathscr{X}} (F_n(x) - F_0(x))^2 \, dP_0(x)$ rather than the historically more appealing $\int_{\mathscr{X}} (F_n(x) - F_0(x))^2 \, dF_0(x)$, since the latter integral has not been defined in general. We shall also need generic rrv's $\bar{D}_n$ and $\bar{W}_n$ corresponding to the special case where F_0 is a uniform df on [0, 1] and $F = F_0$.

THEOREM 4. (i) *D_n is a* rrv.

(ii) *If $F = F_0$ and F_0 is dense, then D_n has the same* df *as $\bar{D}_n$.*

(iii) *If $\alpha_3(P_0) = 0$ and $F = F_0$, then D_n is stochastically no larger than $\bar{D}_n$.*

PROOF. The proof of (i) is based on the formula

$$D_n = \max_{1 \leq k \leq n} (E_+, E_-) , \tag{2}$$

where $E_+ = \max_{1 \leq k \leq n} (k/n - F_0(X_{(k)}))$ and $E_- = \max_{1 \leq k \leq n} (F_0{}^*(X_{(k)}) - (k-1)/n)$, where $X_{(1)}, \cdots, X_{(n)}$ is $X_1, \cdots, X_n$ arranged in ascending order, and where $F_0{}^*(x) = \sup_{u<x} F_0(u)$. With (2), the proof of (i) is straightforward but somewhat tedious (cf., Simons (1972)).

Under the assumptions of (ii), $F_0(X_1), \cdots, F_0(X_n)$ are i.i.d. uniform variables, according to Proposition 6. If one uses them to define $\bar{D}_n$, then $D_n = \bar{D}_n$ follows.

Briefly, one shows (iii) by introducing auxiliary rv's $Y_1, \cdots, Y_n$ (See the proof of Proposition 6, part (ii), and the remark following the proof.) and establishing the inequalities $F_0^0(X_k) \leqq G_0(X_k, Y_k) \leqq F_0(X_k)$, $1 \leqq k \leqq n$ (cf., (1)), in such a way that $G_0(X_k, Y_k)$, $1 \leqq k \leqq n$, are i.i.d. uniform variables. (The Y_i's can be chosen to be i.i.d. uniform random variables which are jointly independent of the X_i's, for instance.) If these are used to define $\bar{D}_n$, then the inequality $D_n \leqq \bar{D}_n$ follows from (2). (Note that under the assumptions $F_0^0 = F_0{}^*$.) □

REMARK. Quite clearly, there is a Glivenko–Cantelli theorem for the D_n's under assumptions (ii) or (iii) of Theorem 4. That is, $D_n \to 0$ almost surely (P_0).

THEOREM 5. (i) *W_n is welldefined (The integrand is $\mathscr{B}$-measurable).*

(ii) *If F_0 is dense, then W_n is a* rrv *and can be computed as the Riemann integral* $\int_0^1 (n^{-1} \sum_{k=1}^n I_{[F_0(X_k) \leq t]} - t)^2 \, dt$.

(iii) *If F_0 is dense and $F = F_0$, then W_n is distributed as $\bar{W}_n$.*

For the sake of brevity, we omit the proof. (cf., Simons (1972)).

REMARKS. 1. Standard statistical tables can be used when applying Theorems 4 and 5.

2. Theorem 3 permits one to apply the Kolmogorov and Smirnov goodness of fit tests (as well as many other nonparametric procedures) much more widely than heretofore. What is required is a random sample with observations in a separable metric space. Thus multivariate data and many types of continuous time series data are suitable candidates for such tests.

3. The linear ordering given in the proof of Theorem 3 is constructable so that, in principle, the test statistic D_n is computable. We have not examined this matter carefully, but there is no obstacle for multivariate data at least. It is not essential to have the linear ordering completely specified in order to get an adequate approximation to D_n. Specifically, if one stops with the partition Z_m and treats each of its elements as if it were a point, Z_m will be a linearly ordered space (induced by the indices on the elements of Z_m). One can compute the analog of D_n, say $D_n^{(m)}$, by identifying each observation X_k with that element of Z_m within which it falls. It can be checked, using (2), that $D_n^{(m)} \nearrow D_n$ as $m \to \infty$. A crude (but perhaps adequate) upper bound on the error $D_n - D_n^{(m)}$ is $\sup_{A \in Z_m} P_0(A)$.

Acknowledgments. The author was aided by helpful conversations with Raymond Cannon, Wassily Hoeffding, J. H. B. Kemperman, Balram Rajput and Flavio Rodrigues. He also wishes to thank the referee for comments which led to improvements and clarifications.

REFERENCES

[1] HALMOS, P. (1950). *Measure Theory.* Van Nostrand, Princeton.

[2] KEMPERMAN, J. H. B. (1956). Generalized tolerance limits. *Ann. Math. Statist.* **27** 180–186.

[3] NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

[4] RODRIGUES, F. W. (1972). Some structural relationships between weak convergence of probability measures and convergence in probability. Ph. D. Dissertation, Univ. of North Carolina, Chapel Hill.

[5] SIMONS, G. (1972). Generalized cumulative distribution functions: I. The linear case with applications to nonparametric statistics. Institute of Statistics Mimeo Series No. 835, Department of Statistics, Univ. of North Carolina.

[6] SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** 261–290.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514