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

GENERALIZED CUMULATIVE DISTRIBUTION FUNCTIONS: I
THE LINEAR CASE WITH APPLICATIONS
TO NONPARAMETRIC STATISTICS

Gordon Simons*

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 835

August, 1972

ABSTRACT

We investigate the behavior of (univariate) cumulative distribution functions which are defined on an abstract linearly ordered space. Special emphasis is given to the study of a class of linearly ordered spaces which J.H.B. Kemperman introduced into the subject of nonparametric tolerance regions. Distribution functions on such spaces can be decomposed. Considerable attention is given to applications. In particular, it is shown how a number of nonparametric statistical procedures can be extended to include situations of multivariate and time dependent data.

GENERALIZED CUMULATIVE DISTRIBUTION FUNCTIONS: I. THE LINEAR CASE WITH APPLICATIONS TO NONPARAMETRIC STATISTICS*

by Gordon Simons

1. INTRODUCTION AND SUMMARY. This paper begins a systematic study of (cumulative) distribution functions defined on an abstract linearly ordered space. J.H.B. Kemperman (1956) has already shown (in an abstract setting) the potential such functions have for nonparametric statistical applications. This potential has motivated our study (even before we were aware of Kemperman's work on generalized tolerance limits) and it strongly influences and somewhat limits the topics discussed here. We find that it is possible to extend the range of applicability of those procedures of nonparametric statistics which are based on the so-called "probability integral transformation" to include situations in which the data is multivariate valued.[1/] (It is sufficient that the data be expressible in a separable metric space.)

Kemperman requires his linearly ordered spaces to satisfy a mild but important countability assumption. The implications of this assumption are more far reaching than his paper suggest. Many of the familiar properties of classical (univariate) distribution functions fail without it. The assumption, which is stated with no reference to probabilistic terms, can be expressed completely in such terms. We do not always restrict our

---

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

[1/] We do not wish to suggest that this possibility is entirely a new idea. Indeed, Wald showed how to compute nonparametric tolerance limits for multivariate data in 1943 [11]!

attention to spaces for which the assumption holds.

Finally, it should be mentioned that Flavio Rodrigues (1972) has found a generalized distribution function (defined on a linearized separable metric space) to be a useful theoretical tool in his work. He is concerned with demonstrating new relationships between the concepts of weak convergence and convergence in probability. We shall not discuss his work further except to indicate points of overlap.

2. BASIC CONCEPTS, TERMINOLOGY AND NOTATION. A nonempty space of points $\mathcal{X}$ is *linearly ordered* if there has been an exclusive assignment of one the relationships $x_1 < x_2$ and $x_2 < x_1$ to each distinct pair x_1, x_2 in $\mathcal{X}$ and the following transitive relationship holds: $x_1 < x_2$ and $x_2 < x_3$ imply x_1 and x_3 are distinct and $x_1 < x_3$. Since this is a well-known concept, we shall use without explanation the related order symbols "$\leq$", "$>$", and "$\geq$" and statements such as "x_1 precedes x_2", "x_2 is larger than x_1" and "$A < B$" for subsets A, B of $\mathcal{X}$.

Let $\mathcal{X} = A \cup B$ with $A < B$. We refer to any set such as A (including $\mathcal{X}$ and the empty set $\emptyset$) as an *initial* and any set such as B as a *terminal*. The smallest σ-field $\mathcal{B}$ containing all of the initials (and terminals) of $\mathcal{X}$ is called the *order σ-field* and its members are called *order sets*. If A (B) has a largest (smallest) point x it is said to be *closed* and we write $A = \overleftarrow{x}$ $(B = \overrightarrow{x})$.[2/] A (B) is *open* if B (A) is closed. Of course, A and B may both be open and closed at the same time. Both

[2/] Since we shall have little need to refer to topology, we let convenience take precedence over common topological terminology with this term and the term "open" which follows.

may be neither closed nor open, as may be seen when $\mathcal{X}$ is the set of rational numbers and $A = \{x \leq \sqrt{2}\}$. Likewise, if $\mathcal{X}$ is the set of real numbers, then $\mathcal{X}$ and $\emptyset$ are neither closed nor open. If $\mathcal{X}$ is the non-negative integers, then $\mathcal{X}$ is a closed *terminal* but not a closed *initial*.

A linearly ordered space $\mathcal{X}$ is called a *kappa space* or, for brevity, a κ *space* if each nonempty initial of $\mathcal{X}$ is expressible as a countable union of closed initials and each nonempty terminal of $\mathcal{X}$ is expressible as a countable union of closed terminals.[3/] When $\mathcal{X}$ is a κ space the order σ-field $\mathcal{B}$ is generated by the closed (alternatively the open) initials of $\mathcal{X}$.

Let $(\Omega, \mathcal{F})$ be a measurable space. A mapping X from Ω into the linearly ordered space $\mathcal{X}$ is called a *random variable* (r.v.) (*in* $\mathcal{X}$) if the inverse image of each order set $B \in \mathcal{B}$ is a set in $\mathcal{F}$. This concept includes the usual *real random variable* (r.r.v.).

Let P be a probability measure on $(\Omega, \mathcal{F})$. The function $F(x) = P(X \leq x)$, $x \in \mathcal{X}$, is called the (cumulative) *distribution function* (d.f.) of X. It is well-defined since $\overleftarrow{x} \in \mathcal{B}$ for each $x \in \mathcal{X}$. Further, $F(X)$ is a r.r.v. since $\{x \in \mathcal{X}\colon\ F(x) \leq t\}$ is an initial for each real t. A d.f. is *dense* if its range is dense in the real interval $[0,1]$. A d.f. F is *right continuous at the point* $x \in \mathcal{X}$ if for each $\epsilon > 0$ there exists an open initial A with $x \in A$ such that $F(u) < F(x)+\epsilon$ for each $u \in A$. A d.f. is *right continuous* if it is right continuous at every point $x \in \mathcal{X}$ for which $\overleftarrow{x}$ is a proper subset of $\mathcal{X}$. A d.f. is *discrete* if there exists

---

[3/] This concept was introduced by J. H. B. Kemperman (1956). Actually, he was working with a slightly more general ordering by allowing distinct pairs of points to be equivalent and not order comparable. All of our results are expressible in his greater generality.

a set of positive constants $\{p_m\}$ (necessarily countable) and corresponding (distinct) points $\{x_m\}$ in $\mathcal{X}$ such that

$$F(x) = \sum_{\{m: x_m \le x\}} p_m, \quad x \in \mathcal{X}.$$

3. THE PROBABILITY INTEGRAL TRANSFORMATION. If F is the d.f. of a random variable X, $F(X)$ (for historical reasons) is called the *probability integral transformation*. It is an important fact in the study of nonparametric statistics that $F(X)$ is a uniformly distributed r.r.v. on $[0,1]$ when X is a r.r.v. and F is continuous.

We now state and (for completeness) prove the following theorem due to Kemperman:

Theorem 1. *Let* X *be a r.v. in the* κ *space* $\mathcal{X}$ *with d.f.* F *arising from the probability measure* P. $F(X)$ *is a uniformly distributed r.r.v. on* $[0,1]$ *if, and only if,* $P(X = x) = 0$ *for every* $x \in \mathcal{X}$.

Proof. Let $0 \le t \le 1$, $A = \{x: F(x) \le t\}$ and $B = \mathcal{X} - A$. Then $F(X)$ is distributed as claimed if we can show

$$P(F(X) \le t) = P(X \in A) = t.$$

Suppose $A \ne \emptyset$. Since $\mathcal{X}$ is a κ space, the initial A may be expressed as $\bigcup_{k=1}^{\infty} \overleftarrow{a}_k$ where $a_1 \le a_2 \le \ldots$ all belong to A. Then $t \ge F(a_k) = P(X \in \overleftarrow{a}_k) \nearrow P(X \in A)$ as $k \to \infty$, and hence, $P(X \in A) \le t$ even if $A = \emptyset$. Similarly, if $B \ne \emptyset$, then $B = \bigcup_{k=1}^{\infty} \overrightarrow{b}_k$ where $b_1 \ge b_2 \ge \ldots$ all belong to B and $t < F(b_k) = P(X \le b_k) = P(X < b_k) = 1 - P(X \in \overrightarrow{b}_k) \searrow 1 - P(X \in B) = P(X \in A)$ as $k \to \infty$. Then, even if $B = \emptyset$, $P(X \in A) \ge t$ and the conclusion follows. The converse is immediate. $\square$

There is a simple necessary and sufficient condition for $F(X)$ to be uniformly distributed on [0,1] which does not require χ to be a κ space:

Theorem 2. *Let* X *be a r.v. in* χ *with d.f.* F. $F(X)$ *is a uniformly distributed r.r.v. on* [0,1] *if, and only if,* F *is dense.*

Proof. If F is dense, then for any $x \in \chi$ with $F(x) \in [0,1)$ and $\epsilon > 0$, there exists a $y \in \chi$ for which $F(x) < F(y) \le F(x)+\epsilon$. Thus $F(x) = P(X \le x) \le P(F(X) \le F(x)) \le P(F(X) < F(y)) \le P(x \le y) = F(y) \le F(x)+\epsilon$. Since $\epsilon > 0$ is arbitrary, it follows that $P(F(X) \le t) = t$ for every t in the (dense) range of F. This implies $F(X)$ is uniformly distributed on [0,1]. The converse is immediate. □

4. KAPPA SPACES. In contrast to the distribution functions of real random variables, the d.f. of a r.v. X may not contain all the information it reasonably should about X. To illustrate this, in Section 7, we exhibit a r.v. X and a family of probability measures $\{P_\alpha, \alpha \in [0,1]\}$ which all give rise to the *same* d.f. F for X. $F(x) = 0$ for all $x \in \chi$ except for a particular point $x_o \in \chi$ at which $F(x) = 1$, and yet $P_\alpha(X = x_o) = \alpha$. (Note the curious implication that $F(X)$ is a Bernoulli variable with mean α under P_α.) Theorem 1 makes clear that this could not happen when X is a r.v. *in a* κ *space*. (Consider the case $\alpha = o$.) This leads us the the following elementary result:

Theorem 3. *Let* X *be a r.v. in the* κ *space* χ *with d.f.* F *arising from the probability measure* P. *Then* F *uniquely determines the value of* $P(X \in B)$ *on the order* σ*-field* $\mathcal{B}$.

Proof. Clearly, F determines the value of $P(X \in B)$ on the semi-ring of sets B of the form $\overleftarrow{x}$ or $\overleftarrow{x}-\overleftarrow{y}$ and consequently on the σ-ring generated by the closed initials. When $\mathbf{X}$ is a κ space this σ-ring is $\mathcal{B}$. $\Box$

With the assumption that $\mathbf{X}$ is a κ space, we can establish useful characterizations of dense and discrete distribution functions:

Theorem 4. *Let* X *be a r.v. in a* κ *space* $\mathbf{X}$ *with d.f.* F. *Then*

I. F *is dense if, and only if,* $P(X=x) = 0$ *for* $x \in \mathbf{X}$.

II. F *is discrete if, and only if,* $\sum_{\{m \geq 1\}} P(X=x_m) = 1$ *for some countable set of distinct points* $\{x_m\}$ *in* $\mathbf{X}$, *and then*

$$F(x) = \sum_{\{m: x_m \leq x\}} P(X=x_m), \quad x \in \mathbf{X}.$$

Proof. I is immediate from Theorems 1 and 2. In showing II, we need the following lemma:

Lemma. *Under the Assumption of Theorem 4:*

$$\sup_{x<u} F(x) = P(X<u), \quad \inf_{x>u} P(X<x) = F(u) \quad \text{and} \quad \sup_{x \in \mathbf{X}} F(x) = 1.$$

(The "sup" is zero and the "inf" is one when their index sets are void.)

Now suppose F is discrete so that F is expressible as

$$F(x) = \sum_{\{m: x_m \leq x\}} p_m, \quad x \in \mathbf{X}.$$

From the lemma, we obtain $\sum_{\{m \geq 1\}} p_m \geq 1$ and $P(X=x_m) = F(x_m) - \sup_{x<x_m} F(x) \geq p_m$. Thus

$$1 \leq \sum_{\{m \geq 1\}} p_m \leq \sum_{\{m \geq 1\}} P\{X=x_m\} \leq 1,$$

and II follows (the converse being trivial). $\Box$

Lemma Proof. Let A be the open initial $\{x:x<u\}$ which is expressible (unless $A = \emptyset$) as $\bigcup_{k=1}^{\infty} \overleftarrow{a_k}$ with $a_1 \le a_2 \le \ldots$ all in A. Then

$$\sup_{x<u} F(x) \ge \lim_{k\to\infty} F(a_k) = P(X\in A) = P(X<u).$$

Since $F(x) \le P(X<u)$ for $x<u$ the first equality of the lemma follows, including the case $A = \emptyset$. The other parts of the lemma are shown similarly. □

Corollary 4.1. *Let* F *be a d.f. associated with a r.v.* X *in the* κ *space* $\mathcal{X}$. *Then* F *can be decomposed into dense and discrete parts as*

$$F(x) = \alpha F_1(x) + (1-\alpha) F_2(x) , x\in\mathcal{X},$$

for some unique $\alpha\in[0,1]$, *where* F_1 *(when* $\alpha>0$*) and* F_2 *(when* $\alpha<1$*) are, respectively, dense and discrete distribution functions of* X *under alternative probability measures.*

Proof. Let P (defined on $(\Omega,\mathcal{F})$) be the probability measure which results in the d.f. F for X and let $\{x_m\}$ denote the set of points x for which $P(X=x) > 0$. Further, let $B = [\omega\in\Omega: X(\omega)\in\{x_m\}]$ with complement B' and $\alpha = P(B')$. If $\alpha = 1$, set $P_1 = P$ so that $F_1(x) \equiv P_1(X\le x) = F(x)$. Since $P_1(X=x) = 0$ for all x, F_1 is dense according to Theorem 4, part I. If $\alpha = 0$, set $P_2 = P$ so that $F_2(x) \equiv P_2(X\le x) = F(x)$. Since $\sum_{\{m\ge1\}} P_2(X=x_m) = 1$, F_2 is discrete according to Theorem 4, part II. If $0<\alpha<1$, set $P_1(A) = P(A|B')$ and $P_2(A) = P(A|B)$, $A\in\mathcal{F}$. $P_1(X=x) = 0$ for all x so F_1 is dense, and $\sum_{\{m\ge1\}} P_2(X=x_m) = 1$ so F_2 is discrete. The decomposition follows. □

Corollary 4.2. *The d.f. of a r.v. in a* κ *space is right continuous.*

Proof. Clearly, dense d.f.'s are right continuous. So, in view of the decomposition given in the previous corollary, we may assume that the d.f.

in question is discrete. There are two kinds of points x at which one must show the d.f. F is right continuous: (i) The closed initial $\overleftarrow{x}$ may also be an open initial (expressible as $\{x: x<v\}$). Then, clearly, $F(u) \leq F(x)+\epsilon$ for all $u\epsilon\overleftarrow{x}$. (ii) The closed initial $\overleftarrow{x}$ may not be an open initial nor equal to the entire space X. Then for any $\epsilon > 0$, $F(u) - F(x)$, which is expressible as $\sum_{\{m: x<x_m \leq u\}} P(X=x_m)$, will be less than ϵ for an infinity of $u>x$. □

We digress briefly before proceeding to the next corollary. Let X and Y be linearly ordered spaces and let $\mathsf{X}\times\mathsf{Y}$ denote the space of pairs (x,y), $x\epsilon\mathsf{X}$, $y\epsilon\mathsf{Y}$, linearly ordered (lexicographically) according to: $(x_1,y_1) < (x_2,y_2)$ if $x_1<x_2$ *or* $x_1=x_2$ and $y_1<y_2$. It is easy to check that the order σ-field for $\mathsf{X}\times\mathsf{Y}$ is a sub-σ-field of the product σ-field generated by the order σ-fields of X and Y respectively. Thus, if X and Y are random variables (arising from the same measurable space) in X and Y respectively, then (X,Y) is a r.v. in $\mathsf{X}\times\mathsf{Y}$. Furthermore, $\mathsf{X}\times\mathsf{Y}$ is a κ space if, and only if, X and Y are κ spaces.

Corollary 4.3. *Let* X *and* Y *be random variables in the* κ *spaces* X *and* Y, *respectively, and let* $F((x,y))$ *denote the d.f. of* (X,Y). *If the* d.f. *of* Y *is dense, then* $F((X,Y))$ *is a uniformly distributed r.r.v. on* $[0,1]$.

Proof. According to Theorem 4, $P((X,Y) = (x,y)) \leq P(Y=y) = 0$ for $(x,y) \epsilon \mathsf{X}\times\mathsf{Y}$ (since the d.f. of Y is dense). The conclusion follows from Theorem 1. □

Remarks

1. This corollary is a generalization of a similar result due to Flavio Rodrigues (1972) when working with real random variables. (C.f., D.A.S.

Fraser (1953) page 50, Kemperman (1956).)

2. This corollary is still true if Y is not a κ space; the proof is more involved. We will not need this greater generality.

Corollary 4.4 *If* X *is a r.v. in a* κ *space with d.f.* F, *then* $F(X)$ *is stochastically at least as large as a uniform variable on* $[0,1]$.

Proof. Without loss of generality, we assume that the underlying probability space admits a r.v. such as Y in the previous corollary (for instance, a uniform variable). Then $F(X) \geq F((X,Y))$, a uniform variable on $[0,1]$ by the previous corollary. □

Remark. Likewise $F(X-)$ (where $F(x-) \equiv \sup_{u<x} F(u) = P(X<x)$) is stochastically no larger than a uniform variable on $[0,1]$. Specifically, $F(X-) \leq F((X,Y)) \leq F(X)$.

We now proceed to give a probabilistic characterization of κ spaces. Let $\mathcal{X}$ be a linearly ordered space with order σ-field $\mathcal{B}$ and let Q be a probability measure on $(\mathcal{X},\mathcal{B})$. It will be recalled that $A\in\mathcal{B}$ is called an *atom* (relative to Q) if $Q(A) > 0$ and for each $B\in\mathcal{B}$ with $B\subset A$, either $Q(B) = 0$ or $Q(B) = Q(A)$. We shall call an atom A a *point atom* if there exists a point $x\in A$ for which $Q(\{x\}) = Q(A)$. Note that $\{x\}\in\mathcal{B}$. (An atom, in general, does not have to be a point atom [as we illustrate in Section 7] when one is working with a measurable space in which all of the single point sets are measurable.)

Theorem 5. $\mathcal{X}$ *is a* κ *space if, and only if, every probability measure* Q *on* $(\mathcal{X},\mathcal{B})$ *has at most point atoms.*

Proof. Suppose the probability measure Q has an atom A. Let $A^* = \{x\colon Q(\overleftarrow{x}A) = 0\}$ and $B^* = \{x\colon Q(\overrightarrow{x}A) = 0\}$. A^* is an intial and B^*

a terminal. If χ is a κ space, then $Q(A^*A) = 0$. For when $A^* \neq \emptyset$, it may be expressed as $\bigcup_{k=1}^{\infty} \overleftarrow{a}_k$ with $a_1 \leq a_2 \leq \ldots$ all in A^*, in which case $0 = Q(\overleftarrow{a}_k A) \nearrow Q(A^*A)$ as $k \to \infty$. Similarly, $Q(B^*A) = 0$. Since $Q(A) > 0$, there must exist a point $x \notin A^* \cup B^*$. Then $Q(\overleftarrow{x}A) = Q(\overrightarrow{x}A) = Q(A) > 0$ (since A is an atom), and it follows that $x \in A$ and $Q(\{x\}) = Q(A)$.

Conversely, suppose that χ is *not* a κ space. Because of the symmetry in the definition of a κ space, we assume without loss of generality that there exists a nonempty initial A which can *not* be expressed as a countable union of closed initials. We proceed to define a probability measure Q on $(\chi, \mathcal{B})$ relative to which A is a non-point atom. Let $\mathcal{S}$ be the semi-ring of "intervals", which are expressible as the set-theoretic difference between two initials. For $S \in \mathcal{S}$, let $Q(S) = 1$ if for some $x \in A$, $S \supseteq \overrightarrow{x}A$, and let $Q(S) = 0$ otherwise. Thus $Q(A) = 1$. Q extends to $\mathcal{B}$ (as a probability measure) providing Q is countably additive on $\mathcal{S}$. Therefore, suppose $S_1, S_2, \ldots$ are (nonempty) disjoint sets in $\mathcal{S}$ and $S = \sum_{k=1}^{\infty} S_k \in \mathcal{S}$. The case $Q(S) = 0$ is trivial, so we only need to show that $Q(S_k) = 1$ for exact one index k when $Q(S) = 1$. In view of the definition of Q on $\mathcal{S}$, there can not be two or more such indices. Now $S \supseteq \overrightarrow{x}A$ for some $x \in A$. Therefore, there will be one such index for which $Q(S_k) = 1$ (i.e., for which $S_k \supseteq \overrightarrow{x}A$ for some $x \in A$) unless there is an infinity of S_k in each set $\overrightarrow{x}A$, $x \in A$. But then, by choosing a point $a_{k'}$ from those sets $S_{k'} \subseteq A$, there results the contradiction $A = \bigcup_{k'} \overleftarrow{a}_{k'}$. Thus Q extends to $\mathcal{B}$ and $Q(B) = 0$ or 1 for every $B \in \mathcal{B}$. Hence, A is an atom. If it were a point atom, A would have to be a closed initial and therefore a countable union of closed initials. □

The reader may wish to skip the remainder of this section which discusses (without proofs) the properties of κ spaces. No subsequent references are made.

Suppose X is a κ space. If Y is a nonempty subspace of X, it too is a κ space relative to the ordering given for X.

Suppose X is any linearly ordered space and let $\bar{X}$ denote the class of initials of X. $\bar{X}$ is linearly ordered by *proper* set inclusion. (That is, "initial A is less than initial B" means "A is a proper subset of B".) A probably more useful subclass is the class of initials which are open and / or not closed (or, equally well, the reverse) which we denote by X^*. All of the initials of $\bar{X}$ and X^* are open or closed. X^* completes X in the sense that X is isomorphic to a subset of X^* and X^{**} is isomorphic to X^*. (If X is the rational numbers in their usual ordering, X^* is ismorphic to the reals.) Finally, X, $\bar{X}$ and X^* are all κ spaces if anyone of them is. We do not know whether a κ space can have a cardinality in excess of that for the reals.

5. SOME NONPARAMETRIC STATISTICS. Roughly speaking, whatever one does in nonparametric statistics with real random variables one can do with random variables defined on a κ space. Kemperman has given one example with his paper on generalized tolerance limits. We give other examples as we discuss several (generalized) statistics associated with the names of Kolmogorov, Smirnov, Cramér and von Mises. We find that all of these have the same distributions as their classical predecessors. This is an obvious advantage, for it permits one to use well established tables in the performance of statistical tests.

We assume the readers of this section are reasonably familiar with the classical concepts and results. We proceed with an informal exposition, devoting special attention to "trouble spots".

For the remainder of this section, $X_1,\ldots,X_n$ denote iid random variables in some linearly ordered space $\mathcal{X}$[4/] with common *conjectured* d.f. F. X denotes a r.v. in $\mathcal{X}$ whose d.f. is F.

Let

$$F_n(x) = n^{-1} \sum_{k=1}^{n} I_{[X_k \le x]} , \; x \in \mathcal{X},$$

(where $I_A = 1$ when A occurs and 0 when it does not) and let

$$D_n = \sup_{x \in \mathcal{X}} |F_n(x) - F(x)|.$$

While it is clear that D_n is a real number in $[0,1]$ for each point ω in the underlying probability space, it is not immediately clear that it is a (real) random variable. It is easy to check that $D_n = \max(E_+,E_-)$ where $E_+ = \max\limits_{1\le k\le n} (\frac{k}{n} - F(X_{(k)}))$, $E- = \max\limits_{1\le k\le n} (FX_{(k)}-) - \frac{k-1}{n})$, where $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ are the values of $X_1,\ldots,X_n$ arranged in ascending order ("order statistics") and where $F(x-) \equiv \sup\limits_{u<x} F(u)$ (zero if the index set is void). Since $F(x)$ and $F(x-)$ are non-decreasing functions, $F(X_k)$ and $F(X_k-)$ are real random variables (and consequently each $F(X_{(k)})$ and $F(X_{(k)}-)$ is a r.r.v.). Thus D_n is a r.r.v.

Finally, let $\bar{D}_n$ denote a generic analog of D_n corresponding to uniform variables $X_1,\ldots,X_n$ on $[0,1]$ with F the true d.f. of such

---

[4/] The concepts of independent random variables and identically distributed random variables in a linearly ordered space are the same as for real random variables.

random variables, and let $\bar{E}_+$ and $\bar{E}_-$ be the corresponding values of E_+ and E_-.

Theorem 6.

I. *If* F *is the true d.f. of each* X_k *and if* F *is dense, then* D_n *is distributed as* $\bar{D}_n$.

II. *If* F *is the true d.f. of each* X_k *and* X *is a* κ *space, then* D_n *is stochastically no larger than* $\bar{D}_n$.

Proof of I. According to Theorem 2, $F(X_1),\ldots,F(X_n)$ are iid uniform variables. Identifying these random variables with those used to define $\bar{D}_n$, we obtain $D_n = \bar{D}_n$. □

Proof of II. Briefly, we use the randomization device illustrated in the proof of Corollary 4.4 and the subsequent remark. One obtains $F(X_k-) \le F((X_k,Y_k)) \le F(X_k)$, where (for definiteness) the Y_k's are iid uniform variables independent of the X_k's, and where the d.f. of each (X_k, Y_k), namely $F((x,y))$, is dense. Relative to the (X_k, Y_k)'s, there are random variables corresponding to D_n, E_+ and E_-. Call then $\tilde{D}_n$, $\tilde{E}_+$ and $\tilde{E}_-$. Then $E_+ \le \tilde{E}_+$ and $E_- \le \tilde{E}_-$ so that $D_n \le \tilde{D}_n$, and (by part I) $\tilde{D}_n$ is distributed as $\bar{D}_n$. □

Quite clearly a Glivenko-Cantelli theorem holds (i.e., $D_n \to 0$ a.s. as $n \to \infty$) when F is the true d.f. if F is dense and / or X is a κ space. We omit the proof.

Cramér (1928), von Mises (1931) and Smirnov (1936) initiated considerable interest in another class of statistics -- ones based on a Stieltjes integral. Cramér and von Mises suggested using

$$(1) \quad \int_{\mathsf{X}} (F_n(x) - F(x))^2 \, dK(x)$$

with $\mathcal{X}$ the real line and K a suitable non-decreasing function. Smirnov suggested the modification

$$(2) \quad \int_{\mathcal{X}} (F_n(x) - F(x))^2\, \Psi(F(x))\, dF(x)$$

where again $\mathcal{X}$ is the real line and Ψ, defined on $[0,1]$, is a suitable non-negative (Borel measurable) weight function.

We shall not try to interpret (1) in our context[5], but (2) can be interpreted as a Lebesque-Stieltjes integral when $\mathcal{X}$ is a κ space: According to Theorem 3, F induces a probability measure on $(\mathcal{X},\mathcal{B})$.[6] Alternatively, if F is dense, it determines a probability measure on the σ-field generated by the closed initials and this measure can be extended to $(\mathcal{X},\mathcal{B})$ by completing it, if necessary.

Our primary concern with (2) is when F is dense, in which case it can be interpreted as the expectation of a real random variable for each point ω in the underlying probability space: Let $Y = F(X)$, a uniform variable on $[0,1]$. Then $F_n(X) = n^{-1} \sum_{k=1}^{n} I_{[F(x_k)\le Y]}$ a.s. where each $X_k(\omega)$ is fixed at the value x_k $(k = 1,\ldots,n)$. This is because $P(I_{[x_k \le X]} \neq I_{[F(x_k)\le F(X)]}) \le P(Y = F(x_k)) = 0$. Consequently, (2) can be expressed as the expectation:

---

[5]One might try to extend the definition of the Riemann-Stieltjes integral using the "refinement of partitions" method. See Apostol (1957), page 192. Alternatively, one might define (1) as a Legesque-Stieltjes integral by trying to associate a measure (on the order σ-field) with K. This is not so easy to do as it is when $\mathcal{X}$ is the real line. Apparently, one would need to study the topological structure of $\mathcal{X}$.

---

[6]The integrand of (2) is $\mathcal{B}$-measurable for each point ω in the underlying probability space.

$$\text{(3)} \quad E\left(n^{-1} \sum_{k=1}^{n} I_{[F(x_k) \le Y]} - Y\right)^2 \Psi(Y) \quad (\omega \text{ fixed}).$$

(3) can be used for computational purposes:

The statistic in question, when F is the conjectured dense d.f. (for the X_k's), is equal to

$$\text{(4)} \quad \int_0^1 \left(n^{-1} \sum_{k=1}^{n} I_{[F(X_k) \le t]} - t\right)^2 \Psi(t)\, dt.$$ [7/]

We state without further proof the following theorem:

Theorem 7. *If* F *is dense and* $\int_0^1 \Psi(t)dt<\infty$, *then* (2) *is a r.r.v. which may be computed as* (4). *Further, if* F *is the true d.f. for each* X_k, *then* (2) *has the same distribution as*

$$\int_0^1 \left(n^{-1} \sum_{k=1}^{n} I_{[U_k \le t]} - t\right)^2 \Psi(t)\, dt$$

where $U_1, U_2, \ldots, U_n$ *are iid uniform variables on* $[0,1]$.

(Nowhere have we assumed that $\mathcal{X}$ is a κ space.)

6. LINEARLY ORDERING A METRIC SPACE. Frequently, data represents measurements in a separable metric space. Such spaces suffice for real valued data and multivariate data. A continuous monotoring of a barometer for a fixed period of time produces an observation in the separable metric space of continuous functions on some bounded closed interval.

In this section, we show that a separable metric space $\mathcal{X}$ can be linearly ordered with the following desirable features:

[7/]This must be a r.r.v. (or possibly an extended real random variable) since it is the limit of a sequence of approximating Riemann sums, each one a r.r.v.

(i) The resulting linearly ordered space is a κ space.

(ii) An observation in $\mathcal{X}$ may be interpreted as resulting from a *random variable* X in $\mathcal{X}$.

(iii) The r.v. X has a dense d.f. unless there are points $x \in \mathcal{X}$ with $P(X = x) > 0$.

(iv) Usually, the linear ordering is constructable and not merely existential. Test statistics can be evaluated (or adequately approximated) to permit the performance of nonparametric tests.

(v) There is flexibility in the construction of the linear ordering. Presumably, one might seek an ordering which produces a test with good power against certain specific alternatives.

On the negative side, it may seem unnatural to linearly order multivariate observations. The flexibility referred to in (v) may be interpreted by some as undesirable arbitrariness. However, most statistical tests are based on the size of a real valued test statistic. This has the implicit effect of "linearly ordering" the sample space in the slightly weaker sense that Kemperman refers to in his 1956 paper. That is, two points which are not comparable are considered as equivalent.

By linearly ordering spaces which do not have a natural ordering, we produce a new class of *nonparametric* statistical tests. These should be evaluated on the basis of their operating characteristics.

Let $\mathcal{X}$ be a separable metric space and $\mathcal{F}$ be the Borel field generated by the open set (equivalently, by the open balls). We consider a sequence of progressively refinining partitions $Z_m = \{A_{i_1, i_2 \ldots, i_m}\}$ $(m \geq 1)$ of $\mathcal{X}$ with the following properties:

(a) Each Z_m contains a countable number of non-void elements belonging to $\mathcal{F}$.

(b) sup $\{|A|: A\in Z_m\} \to 0$ as $m \to \infty$, where $|A|$ denoted the diameter of an arbitrary set $A\in F$.

(c) $A_{i_1,\ldots,i_{m-1}} = \bigcup_{\{i_m\}} A_{i_1,\ldots,i_m}$ $(m\geq 2)$.

(d) Each index i_m is a positive integer.

The definition of a separable metric space easily guarantees that such sequences of partitions exist.

The partitions $\{Z_m\}$ determine a unique linear ordering: For each $x\in X$, there exists a unique sequence of indices $i_1,i_2,\ldots$ such that $x\in A_{i_1,\ldots,i_m}$ for each m. Conversely, condition (b) insures that this sequence is unique to x. (Of course, an arbitrary sequence $i_1,i_2,\ldots$ may not correspond to any point x since $\bigcap_{m=1}^{\infty} A_{i_1,\ldots,i_m}$ might be void). The sequences which correspond to points in X can be linearly ordered lexicographically (as one does with the decimal expansions of the real numbers in $[0,1)$). This induces the linear ordering of X that we associate with the partitions $\{Z_m\}$.

This linear ordering procedure makes each element $A_{i_1,\ldots,i_m}$ of Z_m into an "interval" (the set-theoretic difference of two initials) and linearly orders Z_m $(m\geq 1)$. For instance, each point of $A_{1,2}$ precedes each point of $A_{1,3}$.

Let B denote the order σ-field associated with such a linear ordering. We have the following theorem:

Theorem 8. X, *as ordered by* $\{Z_m, m\geq 1\}$, *is a* κ *space and* $B = F$.

Proof that X is a κ space. Briefly, any nonempty non-closed initial (terminal) A can be expressed as the countable union

$$A = \bigcup \{\overleftarrow{x}_B : B\in Z_m \text{ for some } m\geq 1, B\subseteq A\}$$

$(A = \cup\{\vec{x}_B\colon B\in Z_m$ for some $m\geq 1,\ B\subseteq A\})$

where x_B is an arbitrary point in the set B. □

Proof that $\mathcal{B}\subseteq\mathcal{F}$. Since $\mathcal{X}$ is a κ space, it suffices to show that an arbitrary closed initial $\overleftarrow{x}\in\mathcal{F}$. (The closed initials generate $\mathcal{B}$.) But

$$\overleftarrow{x} = \bigcap_{m=1}^{\infty} B_m$$

where B_m is the (countable) union of all sets in Z_m (and consequently in $\mathcal{F}$) which precede the point x or contain x. □

Proof that $\mathcal{F}\subseteq\mathcal{B}$. Briefly, each open ball $B\in\mathcal{F}$ can be expressed as the countable union

$$B = \cup\{C\colon C\in Z_m \text{ for some } m\geq 1,\ C\subseteq B\},$$

and each ("interval") C is in the order σ-field $\mathcal{B}$. □

With this theorem, the claims (i) - (v), mentioned earlier in this section, are largely apparent: Certainly (i) is. Concerning (ii), an observation X in a metric space $\mathcal{X}$ is generally interpreted to refer to a random mapping into the measurable space $(\mathcal{X},\mathcal{F})$. Since $\mathcal{B}=\mathcal{F}$, (ii) holds. (iii) is a consequence of Theorem 4. Concerning (iv), is quite clear the linear ordering is constructable for finite dimensional Euclidean spaces. The requirements for *adequately* evaluating a (nonparametric) test statistic are less than one might initially suspect since the values of $P(B)$, $B\in Z_m$, determine the values of the d.f. of X (associated with the linear ordering) within certain bounds. When m is large, the upper and lower bounds for $F(x)$ will be quite close to each other. From a practical standpoint, one does not need to completely specify the linear ordering. The specification of a few of the partitions Z_m may be sufficient. Finally, property (v) does not require further comment.

Theorems 5 and 8 combine to produce an indirect proof of the following known result: *If* Q *is a probability measure on* $(\mathcal{X},\mathcal{F})$ *where* $\mathcal{X}$ *is a separable metric space with Borel σ-field* $\mathcal{F}$, *then* Q *is non-atomic if, and only if,* $Q(\{x\}) = 0$ *for each* $x \in \mathcal{X}$. Q *has at most point atoms.*

The approach we have used for defining a linear ordering appears in the work of Flavio Rodrigues (1972). His purpose, like ours, is to define a "univariate" distribution function but for a different reason.

7. COUNTER-EXAMPLES AND COMMENTS. In Section 4, we discuss the importance and properties of κ spaces; in this section, we exhibit linearly ordered spaces which are not κ spaces and demonstrate some of the possible consequences.

Let $\mathcal{X}$ be the set of ordinals which are less than or equal to the first uncountable ordinal Ω.[8/] The order σ-field is

$$\mathcal{B} = \{A\colon A \text{ or } A' \subseteq \overleftarrow{x} + \{\Omega\} \text{ for some } x \in \mathcal{X} - \{\Omega\}\}$$

(where A' denotes the complement of A). Let $\{P_\alpha, \alpha \in [0,1]\}$ denote the family of probability measures on $(\mathcal{X},\mathcal{B})$ which are defined as follows:

$$P_\alpha(A) = \begin{cases} 0 & \text{if } A \subseteq \overleftarrow{x} \text{ for some } x \in \mathcal{X} - \{\Omega\} \\ 1-\alpha & \text{if } \Omega \in A' \subseteq \overleftarrow{x} + \{\Omega\} \text{ for some } x \in \mathcal{X} - \{\Omega\} \\ \alpha & \text{if } \Omega \in A \subseteq \overleftarrow{x} + \{\Omega\} \text{ for some } x \in \mathcal{X} - \{\Omega\} \\ 1 & \text{if } \overrightarrow{x} \subseteq A \text{ for some } x \in \mathcal{X} - \{\Omega\}. \end{cases}$$

[8/] This space has provided many counter-examples in general topology. Halmos (1950), page 231, has used the space to exhibit an interesting non-regular measure. M. Bhaskara Rao and K.P.S. Bhaskara Rao (1971) pursue his example further.

(I.e., Mass α is placed on the maximal point Ω and mass $1-\alpha$ is placed on each "interval" of the form $[x,\Omega)$, $x\in\chi-\Omega$.) Finally, let X be the identity map from χ into χ (i.e., $X(\omega) = \omega$).

Under each probability measure P_α, the d.f. for X is *identically* given by

$$F(x) = \begin{cases} 0 & \text{if } x\in\chi - \Omega \\ 1 & \text{if } x = \Omega. \end{cases}$$

Clearly, the d.f. of X fails to identify the values of $P_\alpha(X\in A)$ for certain sets $A\in\mathcal{B}$. This is due to the fact that χ is not a κ space. The initial $\chi - \{\Omega\}$ can not be expressed as a countable union of closed initials.

Under P_α, F(X) is a Bernoulli variable with mean α. When $\alpha = 0$, $P_\alpha(X = x) = 0$ for all $x\in\chi$. This explains why the κ space assumption is included in Theorem 1.

This example can be used to illustrate several other counter-examples associated with our results in sections 4 and 5. For instance, the Glivenko-Cantelli conclusion fails for $\alpha\in[0,1)$: $D_n \to 1-\alpha$ a.s. as $n \to \infty$. By deleting the point Ω from χ and using the probability measure P_0 on the resultant order σ-field, one finds that the corresponding r.v. X has the d.f. $F(x) \equiv 0$.

Some concluding remarks:

1. These examples seem to suggest that linearly ordered spaces which are not κ spaces are so unpleasant that they should be excluded from any theoretical consideration. Such a position does not appear fully justified. For instance, theorems 2 and 7 do not require a κ space--only a dense d.f. It is not difficult to produce dense d.f.'s on non-κ spaces. (Look at the product of the real line and the space χ given above.)

REFERENCES

[1] Apostol, T.M. (1957). *Mathematical Analysis, A Modern Approach to Advanced Calculus*. Addison-Wesley, Reading.

[2] Chung, K.L. (1968). *A Course in Probability Theory*. Harcourt, Brace and World, New York.

[3] Cramér, H. (1928). "On the composition of elementary errors". *Skand. Aktuarietids*. Vol. 11 13-74 and 141-180.

[4] Fraser, D.A.S. (1953). "Nonparametric tolerance regions". *Ann. Math. Statist.* 24 44-55.

[5] Halmos, P. (1950). *Measure Theory*. Van Nostrand, Princeton.

[6] Kemperman, J.H.B. (1956). "Generalized tolerance limits". *Ann. Math. Statist.* 27 180-186.

[7] Rao, M. Bhaskara and Rao, K.P.S. Bhaskara. (1971). "Borel σ-algebra on $[0,\Omega]$". *Manuscripta Math.* 5 195-198.

[8] Rodrigues, F. W. (1972). "Some structural relationships between weak convergence of probability measures and convergence in probability". Ph. D. Dissertation, University of North Carolina, Chapel Hill.

[9] Smirnov, N.V. (1936). "Sur la distribution de ω^2". *C.R. Acad. Sci. Paris*. Vol. 202 449-452.

[10] von Mises. R. (1931). *Wahrscheinlichkeitsrechnung*. Leipzig-Wien.

[11] Wald, A. (1943). "An extension of Wilks' method for setting tolerance limits". *Ann. Math. Statist.* 14 45-55.