ANNALES DE L'I. H. P., SECTION B

MARGRIT GAUGLHOFER
A. T. BHARUCHA-REID

ε-entropy of sets of probability distribution functions and their Fourier-Stieltjes transforms

Annales de l'I. H. P., section B, tome 9, nº 2 (1973), p. 113-144
<http://www.numdam.org/item?id=AIHPB_1973__9_2_113_0>

Ann. Inst. Henri Poincaré,
Vol. IX, n° 2, 1973, p. 113-144.

Section B :
Calcul des Probabilités et Statistique.

ε-Entropy of sets of probability distribution functions and their Fourier-Stieltjes transforms

by

Margrit GAUGLHOFER and A. T. BHARUCHA-REID
Center for Research in Probability, Department of Mathematics
Wayne State University, Detroit, Michigan 48202

SECTION I

INTRODUCTION

1.1. Introduction

At the present time there are three main known notions of entropy: *measure-theoretic* or *probabilistic*, *epsilon* (ε) or *metric*, and *topological*. Historically, measure-theoretic entropy was the first notion of entropy introduced, and has found widespread applications in probability theory, information theory, and physics. For discussions of measure-theoretic entropy we refer to Khinchin [*13*] and Rényi [*23*, Chap. IX].

The concept of ε-entropy (also called metric entropy) was introduced by Kolmogorov [*14*, *15*] in order to classify compact metric spaces according to their massiveness.

The basic definitions are as follows. Let A be a subset of a metric space $\mathfrak{X}$, and let $\varepsilon > 0$ be given. A family $E_1, E_2, \ldots, E_n$ of subsets of $\mathfrak{X}$ is said to be an *ε-covering of* A if the diameter of each E_k does not exceed 2ε and if the sets E_k cover A. For a given $\varepsilon > 0$, the number n depends on the covering family $\{ E_k \}$, but $N_\varepsilon(A) = \min n$ is an invariant of the set A. The logarithm

$$H_\varepsilon(A) = \log_2 N_\varepsilon(A)$$

is the *ε-entropy of* A.

A related notion is that of the ε-capacity of A. Points $x_1, x_2, \ldots, x_m$ of A are said to be *ε-distinguishable* if the distance between each pair of them exceeds ε. The number $M_\varepsilon(A) = \max m$ is an invariant of the set A ; and the logarithm

$$C_\varepsilon(A) = \log_2 M_\varepsilon(A)$$

is the *ε-capacity of* A.

In general, $H_\varepsilon(A)$ and $C_\varepsilon(A)$ increase rapidly to $+\infty$ as $\varepsilon \to 0$; and their asymptotic behavior serves to describe the compact set A.

Kolmogorov and Vituškin (cf. [*28*]) used the notion of ε-entropy in their famous work on Hilbert's 13th problem. At the present time the concept of ε-entropy is being used to study a number of problems in analysis and probability. In the field of analysis we refer in particular to applications in approximation theory (cf., Lorentz [*17*, *18*]). Some applications of ε-entropy in probability theory are briefly summarized in the next subsection.

The newest concept of entropy is that of topological entropy. We refer to Adler, *et al* [*1*] for a discussion of topological entropy and some of its applications.

Within recent years a number of mathematicians have investigated the relationships between the different concepts of entropy. We refer to Goodwyn [*12*], who showed that topological entropy bounds measure-theoretic entropy ; and to Dinaburg [*5*] who studied the relationship between topological entropy and ε-entropy.

1.2. ε-Entropy in Probability Theory

Although the notion of ε-entropy was defined and introduced in connection with a nonprobabilistic problem, in recent years a number of mathematicians have utilized the concept of ε-entropy to study a number of problems in probability theory.

a. Kolmogorov and Tihomirov [*15*] have used ε-entropy to study certain problems in the probabilistic theory of approximate transmission of signals ; and Prosser [*21*] has computed the ε-entropy and ε-capacity of certain time-varying channels (i. e. certain bounded linear operators on a Hilbert space) in communication theory. Compact, Hilbert-Schmidt, and convolution operators were considered.

Posner and Rodemich [*20*] have used ε-entropy to formulate and study certain problems which arise in the study of data compression. Their results are of great interest in the design and analysis of systems for transmitting and receiving communication signals.

b. Let (S, *d*) be a compact metric space, and let C(S) denote the Banach space of continuous real-valued functions on S. Strassen and Dudley [*26*] utilized the notion of ε-entropy to obtain conditions for which the central limit theorem holds for random variables with values in C(S).

c. Chevet [*2*], Chevet, *et al* [*3*] and Dudley [*6*, *7*] have employed ε-entropy to investigate the sample path continuity of certain random functions with values in a separable metric spaces. These papers are of great importance in the theory of random functions since they introduce a new analytic approach (based on ε-entropy) to the study of sample path continuity. We refer also to the papers of Dudley [*8*] and Sudakov [*27*].

1.3. **Summary**

Following the introduction of the notion of ε-entropy by Kolmogorov, a number of mathematicians have been concerned with the computation of the ε-entropy of certain concrete function spaces, and have utilized these results to study certain problems in analysis (cf., Kolmogorov and Tihomirov [*15*], Lorentz [*18*], and Vituškin [*29*]). In this paper we study the ε-entropy of certain sets of probability distribution functions of real-valued random variables, and we also consider the ε-entropy of the sets of their Fourier-Stieltjes transforms (or characteristic functions).

In Section II we give some basic definitions, and present a summary of some properties of ε-entropy and ε-capacity which will be utilized in subsequent sections. Section III is devoted to a study of the ε-entropy of certain sets of truncated probability distribution functions. In Section IV we consider the ε-entropy of the sets of associated Fourier-Stieltjes transforms.

Section II

ε-ENTROPY : BASIC DEFINITIONS AND PROPERTIES

In this section we give some basic definitions, and present a summary of some properties of ε-entropy and ε-capacity of sets in function spaces. For a detailed treatment of the ε-entropy and ε-capacity of sets in function spaces we refer to Kolmogorov and Tihomirov [*15*].

Let A be a nonempty set in a metric space $\mathfrak{X}$.

DEFINITION 1.1. — A collection $\mathfrak{U}$ of sets $U \subset \mathfrak{X}$ is called an *ε-covering of* A if $\bigcup_{U \in \mathfrak{U}} U \supset A$ and $d(U) \leq 2\varepsilon$ for all $U \in \mathfrak{U}$, where $d(U)$ denotes the diameter of the set U.

DEFINITION 1.2. — A set $U \subset \mathfrak{X}$ is called *ε-separated* if any two distinct points of U are at a distance greater than ε from each other.

DEFINITION 1.3. — A set A is said to be *totally bounded* (or *pre-compact*) is for every $\varepsilon > 0$ there exists a finite ε-covering of A.

Remarks: (*i*) A set A is totally bounded if and only if for every $\varepsilon > 0$ every ε-separated set is finite. (*ii*) Clearly, any compact set is totally bounded.

Throughout this paper we restrict our attention to totally bounded sets.

DEFINITION 1.4. — Let $N_\varepsilon(A)$ be the minimal number of sets in an ε-cover of A. Then $H_\varepsilon(A) = \log_2 N_\varepsilon(A)$ is called the *ε-entropy of* A.

DEFINITION 1.5. — Let $M_\varepsilon(A)$ be the maximal number of points in an ε-separated subset of A. Then $C_\varepsilon(A) = \log_2 M_\varepsilon(A)$ is called the *ε-capacity of* A.

The following properties hold:

(P_1) $H_\varepsilon(A)$ and $C_\varepsilon(A)$ are nondecreasing functions as ε decreases.

$$(P_2)\ C_{2\varepsilon}(A) \leq H_\varepsilon(A) \leq C_\varepsilon(A).$$

For a discussion of the above properties we refer to Kolmogorov and Tihomirov [*15*].

Remark: (*iii*) It is easy to show that for any bounded set A in n-dimensional Euclidean space R_n having interior points $C_\varepsilon(A) \sim H_\varepsilon(A) \sim n \log \frac{1}{\varepsilon}$, where $f \sim g$ means $\lim_{\varepsilon \to 0} f(\varepsilon)/g(\varepsilon) = 1$.

The above fact has given rise to the following generalization of dimension.

DEFINITION 1.6. — The *upper metric dimension* of a set A is defined as

$$\overline{dm}(A) = \limsup_{\varepsilon \to 0} H_\varepsilon(A)/\log(1/\varepsilon)\cdot$$

The *metric dimension* of A is defined as

$$dm(A) = \lim_{\varepsilon \to 0} H_\varepsilon(A)/\log(1/\varepsilon).$$

if this limit exists.

We note that for convex, infinite-dimensional subsets A of a Banach space the metric dimension is always equal to $+\infty$, and is, therefore, useless as a measure to distinguish the massiveness of sets of this type.

DÉFINITION 1.7. — The *exponent* (or *order of growth*) *of entropy* r of A is defined as

$$r(A) = \limsup_{\varepsilon \to 0} \frac{\log H_\varepsilon(A)}{\log (1/\varepsilon)}$$

If the lim sup is equal to a limit, then r is often called the *metric order of* A.

DÉFINITION 1.8. — The *logarithmic order of growth of entropy* is defined as

$$\sigma(A) = \limsup_{\varepsilon \to 0} \frac{\log H_\varepsilon(A)}{\log \log (1/\varepsilon)}.$$

If the lim sup is equal to a limit, then $\sigma(A)$ is often called the *functional dimension of* A.

SECTION III

ε-ENTROPY OF SETS OF PROBABILITY DISTRIBUTION FUNCTIONS

3.1. **Introduction**

Let $(\Omega, \mathscr{A}, \mu)$ be a complete probability measure space; and consider the measurable space $(R, \mathscr{B})$, where R is the real line and $\mathscr{B}$ is the σ-algebra of Borel subsets of R. A mapping $x: \Omega \to R$ is said to be a *real-valued random variable* if $\{ \omega : x(\omega) \in B \} \in \mathscr{A}$ for all $B \in \mathscr{B}$. Associated with every random variable is a real-valued function $F(\xi)$ defined as follows:

$$F(\xi) = \mu(\{ \omega : x(\omega) \in (-\infty, \xi) \}). \tag{3.1}$$

The mapping x induces a measure ν on $\mathscr{B}$ in the following manner:

$$\nu(B) = \mu \circ x^{-1}(B). \tag{3.2}$$

It is easy to show that ν is a probability measure; hence $(R, \mathscr{B}, \nu)$ is called the *induced probability space*.

The function $F(\xi)$ defined by (3.1) is called a *distribution function*. The following properties of distributions are well-known (cf., Gnedenko and Kolmogorov [*11*, Chap. 1] or Loève [*16*, Chap. IV]).

(1) F is non-decreasing;

(2) F is left-continuous:

(3) $$\lim_{\xi \to -\infty} F(\xi) = 0, \qquad \lim_{\xi \to \infty} F(\xi) = 1.$$

Properties (1)-(3) imply the following additional property:

(4) The only discontinuities a distribution function can have are jumps; and the set of discontinuities is at most countable.

Let $\mathfrak{F}$ denote the collection of all distribution functions F. We define a metric d on $\mathfrak{F}$ as follows: For $F, G \in \mathfrak{F}$

$$d(F, G) = \inf_h \{ h : F(\xi - h) - h \leq G(\xi) \leq F(\xi + h) + h, \quad \text{for all } \xi \}. \quad (3.3)$$

The metric defined by (3.3) is called the *Lévy metric*, and $(\mathfrak{F}, d)$ is called the *Lévy space*. We state the following basic result, the proof of which is given in Gnedenko and Kolmogorov [*11*, Chap. 2]:

THEOREM 3.1. — *The Lévy space $(\mathfrak{F}, d)$ is a complete separable metric space.*

In Section 3.2 we compute the ε-entropy of the set of truncated distributions $(\mathfrak{F}_N, d)$; and in Section 3.3 we compute the ε-entropy of the set of truncated discrete distributions.

3.2. ε-Entropy of Sets of Truncated Probability Distribution Functions

Let $\mathfrak{F}_N$ denote the set of distribution functions truncated in the following way:

$$\mathfrak{F}_N = \left\{ G : G(\xi) = \left\{ \begin{array}{lll} 0, & \xi \leq -N & \\ F(\xi), & -N < \xi \leq N, & F \in \mathfrak{F} \\ 1, & N < \xi & \end{array} \right\} \right\}. \quad (3.4)$$

Clearly $\mathfrak{F}_N \subset \mathfrak{F}_{N+1} \subset \ldots$; and $\mathfrak{F}_N \uparrow \mathfrak{F}$ as $N \to \infty$.

We first observe that $\mathfrak{F}_N$ with the uniform metric

$$\rho(F, G) = \sup | F(\xi) - G(\xi) |$$

is not totally bounded for any N; hence a finite ε-entropy cannot be obtained for $(\mathfrak{F}_N, \rho)$. In fact, the set $\{ F_k \}$ where

$$F_k(\xi) = \begin{cases} 0 \ , & \xi < 0 \\ \xi 2^k, & 0 \leq \xi < 2^{-k} \\ 1 \ , & 2^{-k} \leq \xi, \quad k = 1, 2, \ldots, \end{cases} \tag{3.5}$$

is a non-finite, ε-separated set for any $\varepsilon < 1/2$. This follows from the fact that for $n > m$,

$$\max_{\xi} | F_n(\xi) - F_m(\xi) | \geq | F_n(2^{-n}) - F_m(2^{-n}) | \\ = | 1 - 2^m . 2^{-n} | = 1 - 2^{-(n-m)} \geq 1/2.$$

We will now show that $\mathfrak{F}_N$ with the Lévy metric d has finite ε-entropy for any $\varepsilon > 0$, and any finite N.

We now estimate $H_\varepsilon(\mathfrak{F}_N, d)$ —the ε-entropy of $(\mathfrak{F}_N, d)$. This computation will be carried out in two stages: we first compute an upper bound for $H_\varepsilon(\mathfrak{F}_N, d)$, and then compute a lower bound for $H_\varepsilon(\mathfrak{F}_N, d)$.

1. **An upper bound.** — An upper bound for $H_\varepsilon(\mathfrak{F}_N, d)$ is given by the logarithm of the number of elements in an ε-covering of $(\mathfrak{F}_N, d)$. Such an ε-covering can be obtained in the following way: Let

$$c_k = k . 2\varepsilon, \quad k = 0, 1, \ldots, [1/2\varepsilon] = n$$
$$c_{n+1} = 1.$$

An element of the covering is a set of distribution functions whose graphs are contained in the set

$$\{ (x, t) \colon \varphi(x) \leq t \leq \psi(x) \}, \tag{3.6}$$

where $\varphi(x)$ is a nondecreasing step function with the following properties:

(*i*) $\varphi(x) = 0$ for $x \leq -N + 2\varepsilon$.

(*ii*) $\varphi(x)$ has jumps of size $c_m - c_l$ $(0 \leq l \leq m \leq n + 1)$ at $-N + 2k\varepsilon$, $k = 1, 2, \ldots$.

(*iii*) $\varphi(x) = 1$ for $x \geq N$.

The step function ψ has a jump at $x - 2\varepsilon$ of the same size (except when the value 1 is reached) as the function φ at x. We call the set (3.6) an *ε-corridor*. A graphical representation of an ε-corridor is given in Figure 1.

The graph of any distribution function is contained in such an ε-corridor; for if $(x_0, F(x_0)) \in Q_{ik}$, then, as x increases, the fact that F is nondecreasing implies that the graph has to leave Q_{ik} and go into either

$$Q_{i+j,k} (j \geq 1) \text{ or } Q_{i+m,k+1} (m \geq 0).$$

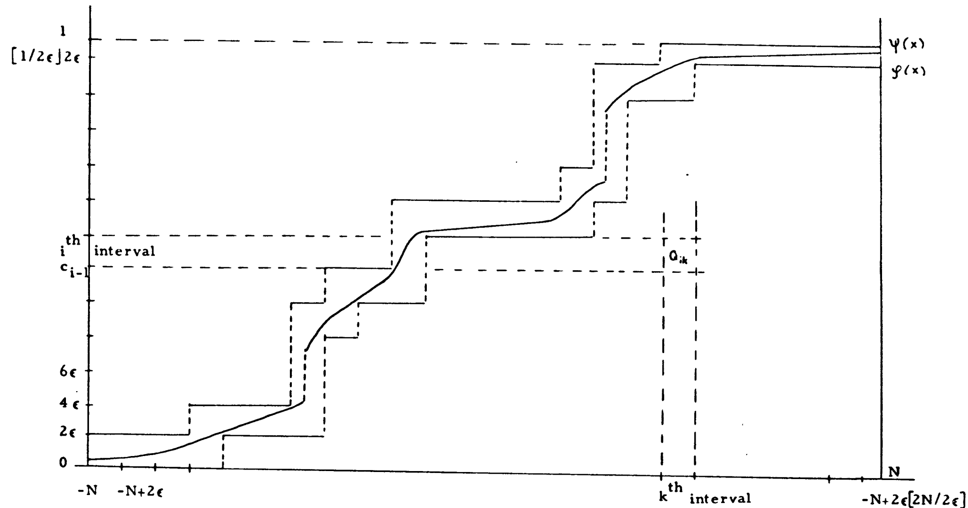

FIG. 1.

The ε-corridor can be extended (or built up) in a manner which depends on the graph of the distribution function F. By the construction of the functions ψ and φ, any two distribution functions F and G belonging to the same ε-corridor have a Lévy distance $d(\mathrm{F}, \mathrm{G}) \leq d(\psi, \varphi) \leq 2\varepsilon$.

We first compute the number of elements in an ε-covering of $(\mathfrak{F}_{\mathrm{N}}, d)$. If the square $\mathrm{Q}_{i-1,k}$, say, is reached by m distinct corridors, and $\mathrm{Q}_{i,k-1}$ by h distinct corridors, then $\mathrm{Q}_{i,k}$ is reached by $m + h$ corridors. Therefore, the entries in the table of Figure 2 indicate the number of distinct corridors

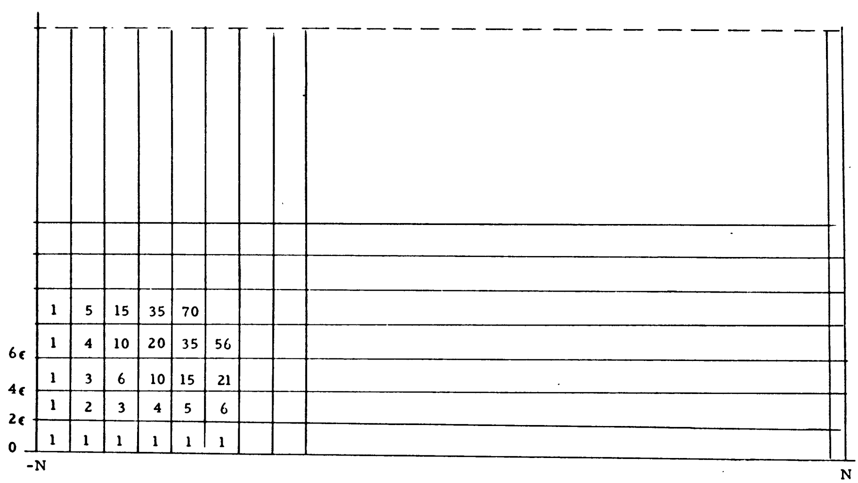

FIG. 2.

reaching the respective square. The sum of the number of corridors reaching the squares $Q_{i,[N/\varepsilon]+1}$, $i = 1, 2, \ldots, [1/2\varepsilon] + 1$ (thats is, the sum of numbers in the last column) is the number of distinct elements in the ε-covering defined above; we remark that if N/ε is an integer, then, here and in what follows, $[N/\varepsilon]$ has to be replaced by $[N/\varepsilon] - 1$; and if $1/2\varepsilon$ is an integer, then $[1/2\varepsilon]$ has to be replaced by $(1/2\varepsilon) - 1$.

Since the number of corridors reaching Q_{ik} is given by the binomial coefficient

$$\binom{(i-1)+(k-1)}{i-1},$$

the ε-covering has

$$S = S(\varepsilon) = \sum_{j=1}^{[1/2\varepsilon]+1} \binom{[N/\varepsilon]+j-1}{j-1} = \sum_{k=0}^{[1/2\varepsilon]} \binom{[N/\varepsilon]+k}{k} \tag{3.7}$$

elements. Hence for any finite N and $\varepsilon > 0$, the number S given by (3.7) is finite. Therefore

$$H_\varepsilon(\mathfrak{F}_N, d) \leq \log S(\varepsilon) \tag{3.8}$$

is finite for any finite N and any $\varepsilon > 0$.

2. **A lower bound.** — It follows from property (P_2) (cf. Sect. II), which expresses the relationship between ε-entropy and ε-capacity, that a lower bound for $H_\varepsilon(\mathfrak{F}_N, d)$ is given by the logarithm of the number of elements in a 2ε-separated subset of $(\mathfrak{F}_N, d)$.

Let $\mathcal{M}(\varepsilon) = \{ F \mid \text{for } x \leq -N + 2\varepsilon[N/\varepsilon],\ F(x) \text{ is equal to the lower boundary function of some } \varepsilon\text{-corridor, while for } x > -N + 2\varepsilon[N/\varepsilon],\ F(x) = 1 \}$.

LEMMA 3.1. — *For any $\varepsilon' > \varepsilon$, $\mathcal{M}(\varepsilon')$ is a 2ε-separated subset of $(\mathfrak{F}_N, d)$.*

Proof. — Let F and G denote two such boundary step functions. In order that they be distinct they must differ on some interval (a, b) of length $2\varepsilon'$ for a value of at least $2\varepsilon'$. Without loss of generality we can take $F(x) < G(x)$ on $(a, b]$. Then $F(x) + \delta < G(x)$ for all $x \in (a, b]$ and all $\delta < 2\varepsilon'$. Since F is constant on $(a, b]$, $F(x) = F(x + \delta)$ as long as x, $x + \delta$ are both contained in $(a, b]$. Put $x = b - \delta$. Then $x \in (a, b]$ and

$$F(x + \delta) + \delta < G(x). \tag{3.9}$$

Now, assume $d(\mathrm{F}, \mathrm{G}) = \delta_0 < 2\varepsilon'$. By definition of the Lévy metric, this implies that for $\delta_0 < \delta < 2\varepsilon'$,

$$\mathrm{F}(x + \delta) + \delta \geq \mathrm{G}(x) \tag{3.10}$$

for all x. But (3.10) is a contradiction to (3.9). Hence we can conclude that $d(\mathrm{F}, \mathrm{G}) \geq 2\varepsilon' > 2\varepsilon$ for any $\varepsilon' > \varepsilon$. Thus the above set of functions is 2ε-separated.

The number of distinct elements in the above $2\varepsilon'$-separated set is equal to

$$\mathrm{R}(\varepsilon') = \sum_{k=0}^{[1/2\varepsilon']} \binom{[\mathrm{N}/\varepsilon'] - 1 + k}{k}.$$

Hence

$$\mathrm{M}_{2\varepsilon}(\mathfrak{F}_{\mathrm{N}}, d) \geq \mathrm{R}(\varepsilon') \text{ (for all } \varepsilon' > \varepsilon).$$

which implies that

$$\mathrm{M}_{2\varepsilon}(\mathfrak{F}_{\mathrm{N}}, d) \geq \mathrm{R}(\varepsilon).$$

Hence

$$\mathrm{H}_{\varepsilon}(\mathfrak{F}_{\mathrm{N}}, d) \geq \mathrm{C}_{2\varepsilon}(\mathfrak{F}_{\mathrm{N}}, d) \geq \log \mathrm{R}(\varepsilon) = \log \sum_{k=0}^{[1/2\varepsilon]} \binom{[\mathrm{N}/\varepsilon] - 1 + k}{k}. \tag{3.11}$$

LEMMA 3.2. $$\sum_{j=0}^{p} \binom{a + j}{j} = \binom{a + p + 1}{p}.$$

Proof. — It is known (cf., Riordan [25], p. 7) that

$$\binom{n}{m} = \sum_{k=0}^{\mathrm{M}} \binom{n - 1 - k}{m - k}$$

where $\mathrm{M} = \min(m, n, -1)$. Therefore,

$$\binom{n}{m} = \sum_{j=0}^{\mathrm{M}} \binom{n - 1 - \mathrm{M} + j}{m - \mathrm{M} + j}.$$

and thus

$$\sum_{j=0}^{p} \binom{a + j}{j} = \sum_{j=0}^{p} \binom{(a + p + 1) - 1 - p + j}{p - p + j}$$
$$= \binom{a + p + 1}{p}.$$

Combining the results obtained above (i. e. (3.8), (3.11) and Lemma 3.2), we can state the following result.

THEOREM 3.2. — *The ε-entropy* $H_\varepsilon(\mathfrak{F}_N, d)$ *of the set of truncated probability distribution functions defined by* (3.4) *satisfies*

$$\log R(\varepsilon) \leq H_\varepsilon(\mathfrak{F}_N, d) \leq \log S(\varepsilon)$$

where

$$S(\varepsilon) = \sum_{k=0}^{[1/2\varepsilon]} \binom{[N/\varepsilon] + k}{k} = \binom{[N/\varepsilon] + [1/2\varepsilon] + 1}{[1/2\varepsilon]}$$

$$R(\varepsilon) = \sum_{k=0}^{[1/2\varepsilon]} \binom{[N/\varepsilon] - 1 + k}{k} = \binom{[N/\varepsilon] + [1/2\varepsilon]}{[1/2\varepsilon]}. \tag{3.12}$$

We now consider the estimation of the exponent, or order of growth, of the entropy of $\mathfrak{F}_N$. Using Stirling's approximation formula

$$n\,! \approx n^{n+\frac{1}{2}}e^{-n}\sqrt{2\pi} \text{ (for large } n)$$

we obtain for p sufficiently large

$$\binom{a + p + 1}{p} = \frac{(a + p + 1)\,!}{p\,!(a + 1)\,!} \approx \frac{(a + p + 1)^{a+p+\frac{3}{2}}}{p^{p+\frac{1}{2}}(a + 1)^{a+\frac{3}{2}}\sqrt{2\pi}}.$$

Therefore

$$\begin{aligned}\log\binom{a + p + 1}{p} &\approx \left(a + p + \frac{3}{2}\right)\log(a + p + 1) - \left(p + \frac{1}{2}\right)\log p \\ &\quad - \left(a + \frac{3}{2}\right)\log(a + 1) - \frac{1}{2}\log(2\pi) \\ &= \left(a + p + \frac{3}{2}\right)\log\left(\frac{a + p + 1}{a + 1}\right) \\ &\quad + p\log\left(\frac{a + 1}{p}\right) - \frac{1}{2}\log(2\pi p).\end{aligned} \tag{3.13}$$

Substituting $a = [N/\varepsilon]$, $p = [1/2\varepsilon]$, we obtain for small ε

$$\begin{aligned}\log S(\varepsilon) \approx \left([N/\varepsilon] + [1/2\varepsilon] + \frac{3}{2}\right)\log\left(1 + \frac{[1/2\varepsilon]}{[N/\varepsilon] + 1}\right) \\ + [1/2\varepsilon]\log\left(\frac{[N/\varepsilon] + 1}{[1/2\varepsilon]}\right) - \frac{1}{2}\log([1/2\varepsilon].2\pi).\end{aligned}$$

Similarly

$$\log R(\varepsilon) = \left([N/\varepsilon] + [1/2\varepsilon] + \frac{1}{2}\right) \log \left(1 + \frac{[1/2\varepsilon]}{[N/\varepsilon]}\right)$$
$$+ [1/2\varepsilon] \log \left(\frac{[N/\varepsilon]}{[1/2\varepsilon]}\right) - \frac{1}{2} \log ([1/2\varepsilon].2\pi).$$

Since $\log (1 + x) \approx x$ for small x, we obtain for large N

$$\left([N/\varepsilon] + [1/2\varepsilon] + \frac{3}{2}\right) \log \left(1 + \frac{[1/2\varepsilon]}{[N/\varepsilon] + 1}\right)$$
$$= O\left(\frac{N}{\varepsilon} \cdot \frac{1}{N}\right) = O\left(\frac{1}{\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0.$$

Similarly

$$\left([N/\varepsilon] + [1/2\varepsilon] + \frac{1}{2}\right) \log \left(1 + \frac{[1/2\varepsilon]}{[N/\varepsilon]}\right) = O\left(\frac{1}{\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0.$$

Both

$$[1/2\varepsilon] \log \left(\frac{[N/\varepsilon] + 1}{[1/2\varepsilon]}\right)$$

and

$$[1/2\varepsilon] \log \left(\frac{[N/\varepsilon]}{[1/2\varepsilon]}\right)$$

are of order

$$O\left(\frac{1}{\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0$$

and

$$O(\log N) \quad \text{as} \quad N \to \infty$$

and

$$\frac{1}{2} \log ([1/2\varepsilon] \cdot 2\pi) = O\left(\log \frac{1}{\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0.$$

Noting that $\log x < x$, we have for any fixed N

$$H_\varepsilon(\mathfrak{F}_N, d) = O\left(\frac{1}{\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0, \tag{3.14}$$

and, for any fixed ε

$$H_\varepsilon(\mathfrak{F}_N, d) = O(\log N) \quad \text{as} \quad N \to \infty. \tag{3.15}$$

Further

$$r = r(\mathfrak{F}_N) = \lim_{\varepsilon \to 0} \frac{\log H_\varepsilon(\mathfrak{F}_N, d)}{\log \dfrac{1}{\varepsilon}} = 1. \tag{3.16}$$

3.3. ε-Entropy of Sets of Truncated Discrete Probability Distribution Functions

Let $\mathscr{D}_N$ denote the class of discrete probability distribution functions truncated at $-N$ and N; i. e., $\mathscr{D}_N = \{ F \in \mathfrak{F}_N \mid F$ is constant except for jumps at $x = k$ of size $p_k \geq 0$, $k = -N, \ldots, N$, $\sum_{-N}^{N} p_k = 1 \}$. It follows from the results of Section 3.2 that the number of ε-corridors necessary to cover $\mathscr{D}_N$ is equal to the number of elements in a 2ε-separated subset. Hence the number of elements is equal to

$$\sum_{k=0}^{[1/2\varepsilon]} \binom{2N + k}{k}$$

Hence we have

THEOREM 3.3. — *The ε-entropy of the set $(\mathscr{D}_N, d)$ of truncated discrete probability distribution functions is given by*

$$H_\varepsilon(\mathscr{D}_N, d) = \log \sum_{k=0}^{[1/2\varepsilon]} \binom{2N + k}{k} = \log \binom{2N + [1/2\varepsilon] + 1}{[1/2\varepsilon]}. \tag{3.17}$$

Next we estimate the order of $H_\varepsilon(\mathscr{D}_N, d)$. If we put $a = 2N$ and $p = [1/2\varepsilon]$ in (3.13) we obtain

$$\begin{aligned} H_\varepsilon(\mathscr{D}_N, d) &= \left(2N + [1/2\varepsilon] + \frac{3}{2}\right) \log (2N + 1 + [1/2\varepsilon]) \\ &\quad - \left([1/2\varepsilon] + \frac{1}{2}\right) \log [1/2\varepsilon] \\ &\quad - \left(2N + \frac{3}{2}\right) \log (2N + 1) - \frac{1}{2} \log 2\pi \\ &= \left(2N + \frac{3}{2}\right) \log \left(\frac{2N + 1 + [1/2\varepsilon]}{2N + 1}\right) \\ &\quad + [1/2\varepsilon] \log \left(\frac{2N + 1 + [1/2\varepsilon]}{[1/2\varepsilon]}\right) - \frac{1}{2} \log ([1/2\varepsilon]2\pi). \end{aligned}$$

Now

$$\begin{aligned} \left(2N + \frac{3}{2}\right) \log \left(1 + \frac{[1/2\varepsilon]}{2N + 1}\right) &= O\left(\log \frac{1}{\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0 \\ &= O(1) \quad \text{as} \quad N \to \infty. \end{aligned}$$

Since $\dfrac{2N+1}{[1/2\varepsilon]}$ is small for ε small

$$[1/2\varepsilon] \log\left(1+\frac{2N+1}{[1/2\varepsilon]}\right) = O(1) \qquad \text{as} \quad \varepsilon \to 0$$

$$= O(\log N) \quad \text{as} \quad N \to \infty$$

$$\frac{1}{2}\log([1/2\varepsilon]\cdot 2\pi) = O\left(\log\frac{1}{\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0.$$

Therefore, for any fixed N

$$H_\varepsilon(\mathscr{D}_N, d) = O\left(\log\frac{1}{\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0 \tag{3.18}$$

and, for any fixed ε

$$H_\varepsilon(\mathscr{D}_N, d) = O(\log N) \quad \text{as} \quad N \to \infty. \tag{3.19}$$

Thus the metric order of $\mathscr{D}_N$ is

$$r(\mathscr{D}_N) = \lim_{\varepsilon\to 0} \frac{\log\log\frac{1}{\varepsilon}}{\log\frac{1}{\varepsilon}} = 0 \tag{3.20}$$

and the logarithmic order is

$$\sigma(\mathscr{D}_N) = \lim_{\varepsilon\to 0} \frac{\log\log\frac{1}{\varepsilon}}{\log\log\frac{1}{\varepsilon}} = 1. \tag{3.21}$$

SECTION IV

ε-ENTROPY OF SETS OF FOURIER-STIELTJES TRANSFORMS OF PROBABILITY DISTRIBUTION FUNCTIONS

4.1. Introduction

Given a probability distribution function $F(\xi)$, the *Fourier-Stieltjes transform* of F, that is

$$f(t) = \int_{-\infty}^{\infty} e^{it\xi} dF(\xi), \qquad t \in R, \tag{4.1}$$

is called the *characteristic function* of F. The following properties of characteristic functions are well-known (cf., Chung [*4*, Chap. VI], Gnedenko and Kolmogorov [*11*, Chap. II], Loève [*16*, Chap. IV], or Lukacs [*19*]):

(1) $|f(t)| \leq f(0) = 1$;

(2) $f(-t) = \overline{f(t)}$;

(3) f is uniformly continuous on the real line;

(4) f is positive-definite.

We remark that to all functions $F + c$, where c is an arbitrary constant, corresponds the same Fourier-Stieltjes transform f. The converse (and, therefore, the 1-1 correspondence between probability distribution functions and characteristic functions) follows from the following inversion formula for Fourier-Stieltjes transforms:

$$F(b) - F(a) = \lim_{u\to\infty} \frac{1}{2\pi} \int_{-u}^{u} \frac{e^{-ita} - e^{-itb}}{it} f(t)dt, \tag{4.2}$$

provided a and b $(a < b)$ are continuity points of F. Also, (4.2) holds for all $a, b \in R$ $(a < b)$, provided F is normalized.

We now state the following basic result (cf., Gnedenko and Kolmogorov [*11*, p. 53]).

THEOREM 4.1. — *Let* F_n, F *denote probability distribution functions and let* f_n, f *denote the associated characteristic functions. Then* F_n *converges to* F *weakly as* $n \to \infty$ (i. e., *the Lévy distance* $d(F_n, F) \to 0$ *as* $n \to \infty$) *if and only if* $f_n(t) \to f(t)$ *uniformly in every bounded interval* $|t| \leq T$.

Let $\mathfrak{F}$ denote the space of all characteristic functions. In view of the above theorem the metrization chosen for $\mathfrak{F}$ is as follows.

DÉFINITION 4.1. — If f and g are two characteristic functions in $\mathfrak{F}$, let

$$\rho(f, g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \| f - g \|_{[-n\pi, n\pi]}, \tag{4.3}$$

where $\| f - g \|_{[a,b]} = \sup_{t\in[a,b]} | f(t) - g(t) |$.

LEMMA 4.1. — $d(F_n, F) \to 0$ *if and only if* $\rho(f_n, f) \to 0$.

Proof. — Assume $\rho(f_n, f) \to 0$. Then to every $\varepsilon > 0$ there exists an N such that $\rho(f_n, f) < \varepsilon$ for all $n > \mathrm{N}$; that is

$$\sum_{m=1}^{\infty} \frac{1}{2^m} \| f_n - f \|_{[-m\pi, m\pi]} < \varepsilon$$

for all $n > \mathrm{N}$. This implies $\| f_n - f \|_{[-m\pi, m\pi]} < \varepsilon \cdot 2^m$ for all $n > \mathrm{N}$. But the last statement says that $f_n \to f$ uniformly in any bounded interval. Therefore, by Theorem 4.1, $d(\mathrm{F}_n, \mathrm{F}) \to 0$ as $n \to \infty$.

Conversely, if $d(\mathrm{F}_n, \mathrm{F}) \to 0$ as $n \to \infty$, then $f_n \to f$ uniformly in every bounded interval $[-m\pi, m\pi]$; that is, for every $\varepsilon > 0$ there exists an $\mathrm{N}(\varepsilon, m)$ such that $\| f_n - f \|_{[-m\pi, m\pi]} < \varepsilon$ for all $n > \mathrm{N}(\varepsilon, m)$, where $\mathrm{N}(\varepsilon, m)$ is nondecreasing as m increases. Now $\| f - g \|_{[-m\pi, m\pi]} \leq 2$ for any m (since $| f(t) - g(t) | \leq | f(t) | + | g(t) | \leq 2$). Given $\varepsilon > 0$, choose M such that

$$\sum_{m=\mathrm{M}+1}^{\infty} \frac{1}{2^m} \cdot 2 < \varepsilon/2.$$

Now, choose N such that $\| f_n - f \|_{[-\mathrm{M}\pi, \mathrm{M}\pi]} < \varepsilon/2$ for all $n > \mathrm{N}$. Consequently, $\| f_n - f \|_{[-m\pi, m\pi]} < \varepsilon/2$ for all $n > \mathrm{N}$ and all $m \leq \mathrm{M}$. Therefore

$$\begin{aligned}
\rho(f_n, f) &= \sum_{m=1}^{\mathrm{M}} \frac{1}{2^m} \| f_n - f \|_{[-m\pi, m\pi]} \\
&\quad + \sum_{m=\mathrm{M}+1}^{\infty} \frac{1}{2^m} \| f_n - f \|_{[-m\pi, m\pi]} \\
&\leq \sum_{m=1}^{\mathrm{M}} \frac{1}{2^m} \cdot \frac{\varepsilon}{2} + \sum_{m=\mathrm{M}+1}^{\infty} \frac{1}{2^m} \cdot 2 \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
\end{aligned}$$

for all $n > \mathrm{N}$.

Let $\tilde{\mathfrak{F}}_{\mathrm{N}}$ denote the space of all characteristic functions corresponding to truncated probability distribution functions $\mathrm{F} \in \mathfrak{F}_{\mathrm{N}}$; and let $\tilde{\mathscr{D}}_{\mathrm{N}}$ denote the space of all characteristic functions corresponding to discrete probability distributions $\mathrm{F} \in \mathscr{D}_{\mathrm{N}}$.

For $\tilde{\mathscr{D}}_{\mathrm{N}}$ we have the following result.

LEMMA 4.2. — *The metric ρ restricted to $\tilde{\mathscr{D}}_{\mathrm{N}}$ coincides with the uniform metric* $\| f - g \| = \sup_{t \in \mathrm{R}} | f(t) - g(t) |$.

Proof. — If $f(t) = \sum_k e^{itk} p_k$, then $f(t + 2\pi) = f(t)$. Therefore

$$\begin{aligned} \| f - g \| &= \sup_{t \in [-\pi,\pi]} | f(t) - g(t) | \\ &= \sup_{t \in [-m\pi,m\pi]} | f(t) - g(t) | \\ &= \| f - g \|_{[-m\pi,m\pi]} \end{aligned}$$

for each $m = 1, 2, \ldots$. Thus

$$\rho(f, g) = \sum_{m=1}^{\infty} \frac{1}{2^m} \| f - g \| = \| f - g \| .$$

It is of interest to note that the fact that two compact metric spaces (X, d) and (Y, ρ) are equivalent (i. e., there is a one-to-one correspondence between X and Y, and $d(x_n, x) \to 0$ if and only if $\rho(y_n, y) \to 0$) does not imply that corresponding sets in X and Y have ε-entropy of same order. We give the following counterexample.

Let $I_n = [0,1]$, $X = \prod_{n=1}^{\infty} I_n$. X, being the product of compact spaces, is compact in the cartesian product topology. Let each I_n be metrized in two ways.

For x_n, y_n in I_n define

$$d_n(x_n, y_n) = \frac{1}{n} | x_n - y_n |$$

and

$$\rho_n(x_n, y_n) = \frac{1}{2^n} | x_n, y_n | .$$

Clearly, for any fixed n, d_n and ρ_n are equivalent metrics on I_n.

Let X be metrized correspondingly: For $x = \{ x_n \}$, $y = \{ y_n \}$ in X define $d(x, y) = \sup_n d_n(x_n, y_n)$ and $\rho(x, y) = \sup_n \rho_n(x_n, y_n)$.

Let $d_n(I_n)$ and $\rho_n(I_n)$ denote the diameter of I_n according to the metrics d_n and ρ_n, respectively. Clearly, $d_n(I_n) \to 0$ and $\rho_n(I_n) \to 0$ as $n \to \infty$. Thus (cf., Dungundji [9], Theorem 7.2, p. 190), d and ρ both metrize the cartesian product topology of the space X, and are therefore equivalent.

Next we compute the ε-entropy of X with respect to both metrics. For $n = 1, 2, \ldots, [1/2\varepsilon]$ let I_n be divided into $[1/2n\varepsilon] + 1$ intervals I_{n,k_n} of length $2n\varepsilon$ (with the exception of the last interval, which may be shorter). Then, for x_n, y_n belonging to the same interval,

$$d_n(x_n, y_n) = \frac{1}{n} | x_n - y_n | \leq \frac{1}{n} \cdot 2n\varepsilon = 2\varepsilon.$$

If $n > [1/2\varepsilon]$, then, for any $x_n, y_n \in \mathrm{I}_n$,

$$d_n(x_n, y_n) \leq \frac{1}{n} \cdot 1 \leq 2\varepsilon.$$

Thus, for x and y in $\displaystyle\prod_{n=1}^{[1/2\varepsilon]} \mathrm{I}_{n,k_n} \times \prod_{[1/2\varepsilon]+1}^{\infty} \mathrm{I}_n$, $d(x, y) = \sup d_n(x_n, y_n) \leq 2\varepsilon$.

Therefore, the family

$$\mathcal{M} = \left\{ \prod_{1}^{[1/2\varepsilon]} \mathrm{I}_{n,k_n} \times \prod_{[1/2\varepsilon]+1}^{\infty} \mathrm{I}_n ; \right.$$
$$\left. k_n = 1, 2, \ldots, [1/2n\varepsilon] + 1 ; n = 1, 2, \ldots, [1/2\varepsilon] \right\}$$

is an ε-covering of X with cardinality

$$\mathrm{N}_\varepsilon = ([1/2\varepsilon] + 1)([1/2 \cdot 2\varepsilon] + 1) \cdots ([1/2[1/2\varepsilon]\varepsilon] + 1).$$

Let

$$\mathrm{S} = \{ (l_{1,k_1}, l_{2,k_2}, \ldots, l_{n,k_n}, 0, \ldots), k_n = 1, 2, \ldots, [1/2n\varepsilon] + 1 ;$$
$$n = 1, 2, \ldots, [1/2\varepsilon] \}$$

where l_{m,k_m} is the left-hand end-point of some interval I_{m,k_m}. Two distinct points of S are at a distance $\geq 2\varepsilon$, since

$$\frac{1}{n} | l_{n_i} - l_{n_j} | \geq \frac{1}{n} \cdot 2n\varepsilon = 2\varepsilon ;$$

thus $\sup_n d_n(l_{n_i} - l_{n_j}) \geq 2\varepsilon$. Therefore, for any $\varepsilon' < \varepsilon$, S is a $2\varepsilon'$-separated set of X. Clearly, S is of the same cardinality as $\mathcal{M}$.

Now,

$$\mathrm{N}_\varepsilon \approx \frac{1}{2\varepsilon} \cdot \frac{1}{4\varepsilon} \cdot \frac{1}{6\varepsilon} \cdots \frac{1}{2[1/2\varepsilon]\varepsilon} \approx \left(\frac{1}{2\varepsilon}\right)^{1/2\varepsilon} \cdot \frac{1}{[1/2\varepsilon]!}.$$

Using Stirling's approximation formula,

$$\mathrm{N}_\varepsilon \approx \left(\frac{\pi}{\varepsilon}\right)^{-1/2} \cdot e^{1/2\varepsilon}.$$

Therefore

$$H_\varepsilon(X, d) \approx -\frac{1}{2}\log\frac{\pi}{\varepsilon} + \frac{1}{2\varepsilon}\log e = \circ\left(\frac{1}{\varepsilon}\right).$$

Similarly, the order of $H_\varepsilon(X, \rho)$ can be obtained. In this case, for $n = 1, 2, \ldots, [\log 1/2\varepsilon]$, the intervals I_n are divided into $[1/2\varepsilon \cdot 2^n] + 1$ intervals I_{n,k_n} of length $2\varepsilon \cdot 2^n$ (or less for the last one).

The sets $\mathscr{M}'$ and S′ (obtained in an analogous way as above) have cardinality

$$N'_\varepsilon \approx \frac{1}{2\varepsilon\cdot 2}\,\frac{1}{2\varepsilon\cdot 2^2}\cdots\frac{1}{2\varepsilon\cdot 2^p} = \left(\frac{1}{2\varepsilon}\right)^p \cdot \frac{1}{2^{1+2+\cdots+p}} = \left(\frac{1}{2\varepsilon}\right)^p\left(\frac{1}{2}\right)^{p(p+1)/2},$$

where $p = [\log 1/2\varepsilon]$. Therefore

$$H_\varepsilon(X, \rho) = p \cdot \log\frac{1}{2\varepsilon} - \left(\frac{p^2}{2} + \frac{p}{2}\right) \approx \frac{p^2}{2} - \frac{p}{2},$$

and

$$H_\varepsilon(X, \rho) = O\left(\left(\log\frac{1}{\varepsilon}\right)^2\right).$$

Hence $H_\varepsilon(X, d)$ and $H_\varepsilon(X, \rho)$ are of different order.

In view of the above, it is necessary to consider, besides $(\mathfrak{F}_N, d)$ and $(\mathscr{D}_N, d)$, also the spaces $(\tilde{\mathfrak{F}}_N, \rho)$ and $(\tilde{\mathscr{D}}_N, \rho)$.

In Section 4.2 we give an estimate of the order of the ε-entropy of $(\tilde{\mathfrak{F}}_N, \rho)$, in Section 4.3 we compute bounds for the ε-entropy of $(\tilde{\mathscr{D}}_N, \rho)$, and in Section 4.4 we estimate the order of the bounds on $H_\varepsilon(\tilde{\mathscr{D}}_N, \rho)$. Finally, in Section 4.5 we present some comments on the entropies $H_\varepsilon(\mathfrak{F}_N, d)$ and $H_\varepsilon(\tilde{\mathfrak{F}}_N, \rho)$.

4.2. An Estimate of the Order of the ε-Entropy of $(\tilde{\mathfrak{F}}_N, \rho)$

Any $f \in \tilde{\mathfrak{F}}_N$ has a representation

$$f(t) = \int_{-N}^{N} e^{itx}dF(x).$$

Put

$$f(z) = \int_{-N}^{N} e^{izx}dF(x), \tag{4.1}$$

where z is complex. It is known (cf., Lukacs [*19*], p. 137, Theorem 7.2.1) that $f(z)$ is an entire functions; also

$$\begin{aligned} |f(z)| &\leq \int_{-N}^{N} |e^{izx}| \, dF(x) \\ &\leq \int_{-N}^{N} e^{|\mathrm{Im}(z)||x|} dF(x) \leq e^{N|\mathrm{Im}(z)|} \leq e^{N|z|}. \end{aligned} \tag{4.2}$$

Using the notation of Kolmogorov and Tihomirov ([*15*], p. 330), $\mathfrak{F}_N$ is contained in $\Phi^1_{1,N}$—the class of entire functions satisfying $|f(z)| \leq e^{N|z|}$. If $\Phi^1_{1,N}$ is considered with the metric $\|f\| = \max\limits_{z \in \bar{K}} |f(z)|$, where $K = \{ z \,|\, |z| < 1 \}$, then it is known (cf., Kolmogorov and Tihomirov [*15*], Theorem 20, p. 330) that

$$C_\varepsilon(\Phi^1_{1,N}) \sim H_\varepsilon(\Phi^1_{1,N}) \sim \frac{\left(\log \frac{1}{\varepsilon}\right)^2}{\log \log \frac{1}{\varepsilon}}. \tag{4.3}$$

We remark that Erohin [*10*] has shown that the above result is not changed if K is replaced by an arbitrary continuum in the z-plane.

We now prove the following result.

THEOREM 4.2. — $$H_\varepsilon(\mathfrak{F}_N, \rho) \lesssim \frac{\left(\log \frac{1}{\varepsilon}\right)^2}{\log \log \frac{1}{\varepsilon}},$$

where $f(\varepsilon) \lesssim g(\varepsilon)$ if $\overline{\lim} \dfrac{f(\varepsilon)}{g(\varepsilon)} \leq 1$.

Proof. — Clearly, $H_\varepsilon(\mathfrak{F}_N, \rho) \leq H_\varepsilon(\Phi^1_{1,N}, \rho)$. We will now show that

$$H_\varepsilon(\Phi^1_{1,N}, \rho) \sim \frac{\left(\log \frac{1}{\varepsilon}\right)^2}{\log \log \frac{1}{\varepsilon}}, \tag{4.4}$$

Let ρ_m denote the metric

$$\rho_m(f, g) = \sup_{-m\pi \leq t \leq m\pi} |f(t) - g(t)|.$$

Since the interval $[-m\pi, m\pi]$ is compact and connected, it is a continuum in the z-plane, and we have

$$C_\varepsilon(\Phi^1_{1,N}, \rho_m) \sim H_\varepsilon(\Phi^1_{1,N}, \rho_m) \sim \frac{\left(\log \frac{1}{\varepsilon}\right)^2}{\log \log \frac{1}{\varepsilon}} \tag{4.5}$$

for any m.

Now

$$\sum_{m=1}^{n} \frac{1}{2^m} \rho_m(f, g) \leq \rho_n(f, g) \sum_{1}^{\infty} \frac{1}{2^m}$$
$$= \rho_n(f, g)$$

for any n. Thus, a minimal ε-cover for $(\Phi^1_{1,N}, \rho_n)$ is also an ε-cover for $\left(\Phi^1_{1,N}, \sum_{1}^{n} \frac{1}{2^m} \rho_m\right)$, and therefore

$$H_\varepsilon\left(\Phi^1_{1,N}, \sum_{1}^{n} \frac{1}{2^m} \rho_m\right) \leq H_\varepsilon(\Phi^1_{1,N}, \rho_n) \sim \frac{\left(\log \frac{1}{\varepsilon}\right)^2}{\log \log \frac{1}{\varepsilon}}$$

for every n. Letting $n \to \infty$, we obtain

$$H_\varepsilon(\Phi^1_{1,N}, \rho) \overset{\infty}{\lesssim} \frac{\left(\log \frac{1}{\varepsilon}\right)^2}{\log \log \frac{1}{\varepsilon}}. \tag{4.6}$$

On the other hand, since

$$\rho(f, g) = \sum_{m=1}^{\infty} \frac{1}{2^m} \rho_m(f, g) \geq \rho_1(f, g) \cdot \sum_{1}^{\infty} \frac{1}{2^m} = \rho_1(f, g),$$

a maximal 2ε-separated set in $(\Phi^1_{1,N}, \rho_1)$ is also a 2ε-separated set in $(\Phi^1_{1,N}, \rho)$; and therefore

$$H_\varepsilon(\Phi^1_{1,N}, \rho) \geq C_{2\varepsilon}(\Phi^1_{1,N}, \rho) \geq C_{2\varepsilon}(\Phi^1_{1,N}, \rho_1) \sim \frac{\left(\log \frac{1}{\varepsilon}\right)^2}{\log \log \frac{1}{\varepsilon}};$$

i. e.

$$H_\varepsilon(\Phi^1_{1,N}, \rho) \precsim \frac{\left(\log \frac{1}{\varepsilon}\right)^2}{\log \log \frac{1}{\varepsilon}}. \tag{4.7}$$

(4.6) and (4.7) together imply (4.4), and thus the theorem is proved.

4.3. Bounds for the ε-Entropy of $(\tilde{\mathscr{D}}_N, \rho)$

In order to find an upper bound for $H_\varepsilon(\tilde{\mathscr{D}}_N, \rho)$, we construct an ε-covering of $\tilde{\mathscr{D}}_N$, the class of characteristic functions corresponding to the discrete distributions in $\mathscr{D}_N$. Given $f \in \tilde{\mathscr{D}}_N$, f has a representation

$$f(t) = \sum_{k=-N}^{N} e^{itk} p_k, \tag{4.8}$$

where $\sum_{k=-N}^{N} p_k = 1$. Now f induces a vector $(m_{-N}, \ldots, m_0, \ldots, m_N)$ of positive odd integers in the following way: let

$$m_k = \begin{cases} \left[\dfrac{(2N+1)p_k}{\varepsilon}\right], & \text{if } \left[\dfrac{(2N+1)p_k}{\varepsilon}\right] \text{ is odd} \\ \left[\dfrac{(2N+1)p_k}{\varepsilon}\right] + 1, & \text{if } \left[\dfrac{(2N+1)p_k}{\varepsilon}\right] \text{ is even.} \end{cases}$$

Let $U((m_k)_{k=-N}^N)$ be the set of all functions $f \in \tilde{\mathscr{D}}_N$ inducing the same vector $(m_k)_{k=-N}^N$. Clearly, any function $f \in \tilde{\mathscr{D}}_N$ is in some $U((m_k)_{k=-N}^N)$.

Any two functions belonging to the same set $U((m_k)_{k=-N}^N)$ are at a distance less than or equal to 2ε. This can be verified as follows. If $f, g \in U((m_k)_{k=-N}^N)$, then

$$\begin{aligned} |f(t) - g(t)| &= \left| f(t) - \sum_{k=-N}^{N} e^{itk} m_k \cdot \frac{\varepsilon}{2N+1} + \sum_{k=-N}^{N} e^{itk} m_k \cdot \frac{\varepsilon}{2N+1} - g(t) \right| \\ &\leq \left| f(t) - \sum_{k=-N}^{N} e^{itk} m_k \cdot \frac{\varepsilon}{2N+1} \right| \\ &\quad + \left| g(t) - \sum_{k=-N}^{N} e^{itk} m_k \cdot \frac{\varepsilon}{2N+1} \right|. \end{aligned}$$

But

$$\left| f(t) - \sum_{k=-N}^{N} e^{itk} m_k \cdot \frac{\varepsilon}{2N+1} \right| \leq \sum_{k=-N}^{N} | e^{itk} | \left| p_k - \frac{m_k \cdot \varepsilon}{2N+1} \right|$$

$$= \sum_{k=-N}^{N} \frac{\varepsilon}{2N+1} \left| p_k \cdot \frac{(2N+1)}{\varepsilon} - m_k \right|$$

$$\leq \sum_{k=-N}^{N} \frac{\varepsilon}{2N+1} \cdot 1 = \varepsilon.$$

Similarly, $\left| g(t) - \sum_{k=-N}^{N} e^{itk} m_k \cdot \frac{\varepsilon}{2N+1} \right| \leq \varepsilon$. Therefore, $| f(t) - g(t) | \leq 2\varepsilon$, for all t, thus $\rho(f, g) \leq 2\varepsilon$. Thus, the class of sets $U\big((m_k)_{k=-N}^{N}\big)$ is an ε-covering of $(\tilde{\mathscr{D}}_N, \rho)$. The number of sets in it is equal to the number of different vectors $(m_k)_{k=-N}^{N}$ induced by all the characteristic functions of $\tilde{\mathscr{D}}_N$. To count their numbers, we introduce the following abbreviations: $\delta = \frac{\varepsilon}{2N+1}$, $2M + 1 = \left[\frac{2N+1}{\varepsilon}\right]$ or $\left[\frac{2N+1}{\varepsilon}\right] + 1$ (whichever is odd), $m = \left[\frac{1}{2\varepsilon}\right]$. Now,

$$m_k = \begin{cases} 1 & \text{for} \quad 0 \leq p_k < 2\delta \\ 3 & \text{for} \quad 2\delta \leq p_k < 4\delta \\ \vdots & \\ 2M+1 & \text{for} \quad 2M\delta \leq p_k \leq 1. \end{cases}$$

LEMMA 4.3. — $\sum_{k=-N}^{N} m_k$ *is an odd integer between* $s_0 = 2(m + N) + 1$ *and* $s_1 = 2(M + N) + 1$.

Proof. — Clearly $\sum_{k=-N}^{N} m_k$, as a sum of an odd number of odd integers, is odd. Since $\sum_{k=-N}^{N} p_k = 1$, $(2N + 1) \max_{-N \leq k \leq N} p_k \geq 1$, therefore

$$\max_{-N \leq k \leq N} p_k \geq \frac{1}{2N+1} = 2\delta \cdot \frac{1}{2\varepsilon}.$$

Thus one of the $m_k's$ is at least $2 \cdot \left[\frac{1}{2\varepsilon}\right] + 1$; the other $m_k's$ at least 1, so $\sum_{k=-N}^{N} m_k \geq s_0 = 2(m + N) + 1$. The maximum value any of the $m_k's$ can take on is $2M + 1$; if this value is taken on, all other $m_k's$ must be 1, so $\sum_{k=-N}^{N} m_k \leq s_1 = 2(M + N) + 1$.

We now count the number of vectors $(m_k)_{k=-N}^{N}$ which lead to a sum $\sum_{k=-N}^{N} m_k = 2s+1$, then take the sum with respect to s, $m+N \leq s \leq M+N$.

LEMMA 4.4. — *There are* $\binom{N+s}{2N}$ *distinct vectors* $(m_k)_{k=-N}^{N}$ *whose components add up to* $2s + 1$; i. e. *there are* $\binom{N+s}{2N}$ *compositions of* $2s + 1$ *with* $2N + 1$ *odd parts.*

Proof. — The generating function for compositions with m odd parts is

$$f_m(t) = (t + t^3 + t^5 + \cdots)^m = \left(\frac{t}{1 - t^2}\right)^m = \Sigma c_{m,n} t^n,$$

where the coefficient $c_{m,n}$ of t^n is the number of compositions of n with m odd parts (cf., Riordan [*24*], p. 124). Now,

$$(1 - t^2)^{-m} = \sum_{r=0}^{\infty} \binom{m + r - 1}{r} t^{2r},$$

therefore

$$f_m(t) = \sum_{r=0}^{\infty} \binom{m + r - 1}{m - 1} t^{m+2r}.$$

For $m = 2N + 1$, the coefficient of t^{2s+1} is

$$\binom{2N + 1 + s - N - 1}{2N} = \binom{N + s}{2N}.$$

This completes the proof of Lemma 4.4.

The total number of distinct induced vectors $(m_k)_{k=-N}^{N}$, and therefore the number of elements in the ε-covering defined above, is

$$R = \sum_{s=m+N}^{M+N} \binom{N+s}{2N} = \sum_{k=m}^{M} \binom{2N+k}{2N} = \sum_{k=m}^{M} \binom{2N+k}{k}. \tag{4.9}$$

From the above and an application of Lemma 3.2, we have

$$\begin{aligned} H_\varepsilon(\tilde{\mathscr{D}}_N, \rho) &\le \log\left(\sum_{k=m}^{M} \binom{2N+k}{k}\right) \\ &= \log\left(\binom{2N+M+1}{M} - \binom{2N+m}{m-1}\right), \end{aligned} \tag{4.10}$$

where $m = [1/2\varepsilon]$, $M = \left[\frac{1}{2}\left[\frac{2N+1}{\varepsilon}\right]\right]$.

In order to find a *lower bound* for $H_\varepsilon(\tilde{\mathscr{D}}_N, \rho)$, we construct a 2ε-separated subset of $\tilde{\mathscr{D}}_N$. An element $f(t) = \sum_{k=-N}^{N} e^{itk} p_k$ is defined by a vector $(p_k)_{k=-N}^{N}$. Consider the set

$$D = \left\{ (p_k)_{k=-N}^{N} \mid p_k = r_k \cdot \frac{1}{p}, \quad r_k \text{ integers}, \quad \Sigma r_k = p, \quad 1 \le p \le \sqrt{m} \right\},$$

where $m = [1/2\varepsilon]$.

We now show that two elements f, g defined by distinct vectors $(p_k)_{k=-N}^{N}$ and $(q_k)_{k=-N}^{N}$ in D, respectively, are at a distance greater than or equal to 2ε. Let $p_k = r_k \cdot \frac{1}{p}$, $q_k = s_k \cdot \frac{1}{q}$. Since the two vectors are distinct, for at least one k, $p_k \neq q_k$, which implies that $r_k q \neq s_k p$, and, since r_k, s_k, p and q are integers, $|r_k q - s_k p| \ge 1$. Let $r_k q - s_k p = d_k$, $d_{-N+k} = a_k$. Then $p_k - q_k = \frac{1}{pq} \cdot d_k$.

We require an estimate of $\rho(f, g)$, where

$$\begin{aligned} \rho(f, g) &= \sup_t \left| \sum_{k=-N}^{N} e^{itk}(p_k - q_k) \right| = \frac{1}{pq} \sup_t \left| \sum_{k=-N}^{N} e^{itk} d_k \right| \\ &= \frac{1}{pq} \sup_t \left| \sum_{k=0}^{2N+1} e^{itk} a_k \right|. \end{aligned}$$

Let $e^{it} = z$.

LEMMA 4.5. — *Let a_k, $k = 0, \ldots, n$ be integers, $a_k \neq 0$ for at least one k. Then*

$$\max_{|z|=1} \left| \sum_{k=0}^{n} a_k z^k \right| \geq 1.$$

Proof. — Let $w(z) = \sum_{0}^{n} a_k z^k$. Since the maximum of an analytic function is assumed on the boundary,

$$\max_{|z|=1} |w(z)| = \max_{|z|\leq 1} |w(z)|.$$

If $a_0 \neq 0$,

$$\max_{|z|\leq 1} |w(z)| \geq |w(0)| = |a_0| \geq 1.$$

If $a_0 = 0$, let i denote the index of the first coefficient which is not equal to zero, i. e. $a_1 = 0, \ldots, a_{i-1} = 0$, $a_i \neq 0$. Then

$$\begin{aligned} \max_{|z|=1} |w(z)| &= \max_{|z|=1} |z^i| |a_i + a_{i+1}z + \cdots + a_n z^{n-i}| \\ &= \max_{|z|\leq 1} |w'(z)| \geq |w'(0)| = |a_i| \geq 1. \end{aligned}$$

Therefore,

$$\rho(f, g) = \frac{1}{pq} \sup_t \left| \sum_{k=0}^{2N+1} e^{itk} a_k \right| \geq \frac{1}{pq} \geq \frac{1}{m} \geq 2\varepsilon.$$

We now determine the number of elements in the 2ε-separated set defined by the vectors in D. We first count the number of vectors $(r_k)_{k=-N}^{N}$ which lead to a sum p, then take the sum with respect to p, $p = 1, 2, \ldots, [\sqrt{m}]$.

LEMMA 4.6. — *There are $\binom{p-1}{r-1}\binom{2N+1}{r}$ vectors $(r_k)_{k=-N}^{N}$ having exactly r components not equal to zero, adding up to p.*

Proof. — There are $\binom{2N+1}{r}$ ways to place the r non-zero numbers in the $2N + 1$ positions. The generating function for compositions with exactly r parts is $g_r(t) = (t + t^2 + \cdots)^r$. Thus the number of compositions of p into r parts is the coefficient of t^p in the expansion of $g_r(t)$, which is equal to $\binom{p-1}{r-1}$ (cf., Riordan [*24*], p. 124). Therefore, the product $\binom{2N+1}{r} \cdot \binom{p-1}{r-1}$ gives the number of vectors with r positive components and sum equal to p.

If we take the sum of the above expression, first with respect to $r \leq \min(p, 2N + 1)$, then with respect to $p \leq [\sqrt{m}]$, we find the total number R of elements in the set D.

In our notation we use the boundary convention

$$\binom{n}{n+m} = 0, \tag{4.11}$$

for $n = 0, 1, \ldots ; m = 1, 2, \ldots$. Therefore

$$R = \sum_{p=1}^{[\sqrt{m}]} \sum_{r=1}^{p} \binom{p-1}{r-1}\binom{2N+1}{r}. \tag{4.12}$$

This expression can be simplified, if use is made of the following identities (cf., Riordan [*25*], p. 3 and p. 9).

$$\binom{a}{b} = \binom{a}{a-b} \tag{4.13}$$

$$\binom{a+b}{c} = \sum_{d=0}^{c} \binom{a}{d}\binom{a}{c-d} \tag{4.14}$$

(4.14), together with the boundary convention (4.12), and (4.13) imply

$$\begin{aligned}\binom{2N+p}{p} &= \sum_{r=0}^{p} \binom{2N+1}{r}\binom{p-1}{p-r} = \sum_{r=1}^{p} \binom{2N+1}{r}\binom{p-1}{p-r} \\ &= \sum_{r=1}^{p} \binom{2N+1}{r}\binom{p-1}{r-1}.\end{aligned}$$

Thus

$$R = \sum_{k=1}^{[\sqrt{m}]} \binom{2N+k}{k}. \tag{4.15}$$

Using the above results and Lemma 3.2, we have

$$\begin{aligned}H_\varepsilon(\tilde{\mathscr{D}}_N, \rho) &\geq \log\left(\sum_{k=1}^{[\sqrt{m}]} \binom{2N+k}{k}\right) \\ &= \log\left(\binom{2N+[\sqrt{m}]+1}{[\sqrt{m}]} - 1\right).\end{aligned} \tag{4.16}$$

We now utilise the following result of Kolmogorov and Tihomirov ([15], Theorem XXI, p. 331) to obtain an estimate of the order of $H_\varepsilon(\tilde{\mathscr{D}}_N, \rho)$. Let $F^1_{p,N}$ denote the class of all entire, 2π-periodic functions $f(z)$ with $|f(z)| \leq e^{N|\mathrm{Im}(z)|^p}$, $p > 1$. Then

$$C_\varepsilon(F^1_{p,N}) \sim H_\varepsilon(F^1_{p,N}) \sim \frac{4\sigma^{\frac{1}{p}}p^2}{(2p-1)[(p-1)\log e]^{1-1/p}}\left(\log\frac{1}{\varepsilon}\right)^{2-\frac{1}{p}}, \tag{4.17}$$

where $F^1_{p,N}$ is considered in the uniform metric.

If we let $p \downarrow 1$, and note that

$$\lim_{p\downarrow 1}(p-1)^{(p-1)/p} = 1,$$

then the above result can be extended to $p = 1$; that is

$$C_\varepsilon(F^1_{1,N}) \sim H_\varepsilon(F^1_{1,N}) \sim 4N\log\frac{1}{\varepsilon}. \tag{4.18}$$

Now, for $f \in \tilde{\mathscr{D}}_N$, $f(t)$ has a representation given by (4.8). Define

$$f(z) = \sum_{-N}^{N} e^{izk}p_k.$$

Clearly $f(z)$ is an entire functions; and

$$|f(z)| \leq \sum_{k=-N}^{N}|e^{izk}|p_k \leq e^{|\mathrm{Im}(z)|N}\sum_{k=-N}^{N}p_k = e^{|\mathrm{Im}(z)|N},$$

$$f(z+2\pi) = \sum_{k=-N}^{N}e^{i(z+2\pi)k}p_k = \sum_{k=-N}^{N}e^{izk}e^{i2\pi k}p_k = \sum_{k=-N}^{N}e^{izk}p_k = f(z).$$

Therefore, $\tilde{\mathscr{D}}_N \subset F^1_{1,N}$; and using (4.18) we have

$$H_\varepsilon(\tilde{\mathscr{D}}_N, \rho) \sim 4N\log\frac{1}{\varepsilon}. \tag{4.19}$$

4.4. The Estimation of the Order of the Bounds of $H_\varepsilon(\tilde{\mathscr{D}}_N, \rho)$

In the last section an upper bound of $H_\varepsilon(\tilde{\mathscr{D}}_N, \rho)$ was found to be

$$\log\left(\binom{2N+M+1}{M} - \binom{2N+m}{m-1}\right),$$

where

$$m = \left[\frac{1}{2\varepsilon}\right], \qquad M = \left[\frac{1}{2}\left[\frac{(2N+1)}{\varepsilon}\right]\right].$$

Now

$$\begin{aligned}
\log\left(\binom{2N+M+1}{M} - \binom{2N+m}{m-1}\right) &\leq \log\binom{2N+M+1}{M} \\
&\approx \left(2N+M+\frac{3}{2}\right)\log(2N+M+1) \\
&- \left(M+\frac{1}{2}\right)\log M - \left(2N+\frac{3}{2}\right)\log(2N+1) - \frac{1}{2}\log 2\pi \\
&= \left(2N+\frac{3}{2}\right)\log\left(1+\frac{M}{2N+1}\right) \\
&+ M\log\left(1+\frac{2N+1}{M}\right) - \frac{1}{2}\log(2\pi M).
\end{aligned}$$

Noting that $M \approx \dfrac{N}{\varepsilon}$, we find that

$$\begin{aligned}
\left(2N+\frac{3}{2}\right)\log\left(1+\frac{M}{2N+1}\right) &= O\left(\log\frac{1}{\varepsilon}\right) && \text{as} \quad \varepsilon \to 0 \\
&= O(N) && \text{as} \quad N \to \infty
\end{aligned}$$

$$\begin{aligned}
M\log\left(1+\frac{2N+1}{M}\right) = O\left(M\cdot\frac{1}{M}\right) &= O(1) && \text{as} \quad \varepsilon \to 0 \\
&= O(N) && \text{as} \quad N \to \infty
\end{aligned}$$

$$\begin{aligned}
\frac{1}{2}\log(2\pi M) &= O\left(\log\frac{1}{\varepsilon}\right) && \text{as} \quad \varepsilon \to 0 \\
&= O(\log N) && \text{as} \quad N \to \infty.
\end{aligned}$$

Thus, the *upper bound* (4.10) of $H_\varepsilon(\tilde{\mathscr{D}}_N, \rho)$ is of order less than or equal to $O\left(\log\frac{1}{\varepsilon}\right)$ as $\varepsilon \to 0$, for any fixed N; and less than or equal to O(N) as $N \to \infty$, for any fixed ε.

A lower bound for $H_\varepsilon(\tilde{\mathscr{D}}_N, \rho)$ is

$$\log\left(\binom{2N+[\sqrt{m}]+1}{[\sqrt{m}]} - 1\right)$$

where $m = [1/2\varepsilon]$. This expression is (for large N) of the same order as $\log \binom{2N + [\sqrt{m}] + 1}{[\sqrt{m}]}$. Now

$$\log \binom{2N + [\sqrt{m}] + 1}{[\sqrt{m}]} \approx \left(2N + \frac{3}{2}\right) \log \left(1 + \frac{[\sqrt{m}]}{2N + 1}\right) + [\sqrt{m}] \log \left(1 + \frac{2N + 1}{[\sqrt{m}]}\right) - \frac{1}{2} \log (2\pi[\sqrt{m}]).$$

Noting that $[\sqrt{m}] \approx (2\varepsilon)^{-1/2}$, and using the fact that $\log x^n = n \log x$, we find that

$$\left(2N + \frac{3}{2}\right) \log \left(1 + \frac{[\sqrt{m}]}{2N + 1}\right) = O(\log \sqrt{1/\varepsilon}) = O\left(\log \frac{1}{\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0$$
$$= O(1) \quad \text{as} \quad N \to \infty.$$

$$[\sqrt{m}] \log \left(1 + \frac{2N + 1}{[\sqrt{m}]}\right) = O(1) \quad \text{as} \quad \varepsilon \to 0$$
$$= O\,(\log N) \quad \text{as} \quad N \to \infty.$$

$$\frac{1}{2} \log (2\pi[\sqrt{m}]) = O\,(\log \sqrt{1/\varepsilon}) = O\left(\log \frac{1}{\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0.$$

Thus, the *lower bound* (4.16) is, for any fixed N, of order $O\left(\log \frac{1}{\varepsilon}\right)$ as $\varepsilon \to 0$; and, for any fixed ε, of order O (log N) as $N \to \infty$.

We can conclude that

$$H_\varepsilon(\tilde{\mathscr{D}}_N, \rho) = O\left(\log \frac{1}{\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0, \tag{4.20}$$

and

$$O\,(\log N) \leq H_\varepsilon(\tilde{\mathscr{D}}_N, \rho) \leq O(N) \quad \text{as} \quad N \to \infty. \tag{4.21}$$

4.5. **Concluding Remarks**

In Section 3.2 we showed that

$$H_\varepsilon(\mathfrak{F}_N, d) = O\left(\frac{1}{\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0;$$

and in Section 4.2 we showed that

$$H_\varepsilon(\tilde{\mathfrak{F}}_N, \rho) \leq O\left(\frac{(\log 1/\varepsilon)^2}{\log \log 1/\varepsilon}\right) \quad \text{as} \quad \varepsilon \to 0.$$

It is of interest to note that the orders of entropy of the two sets are quite different. In view of Lemma 4.1, $(\mathfrak{F}_N, d)$ and $(\tilde{\mathfrak{F}}_N, \rho)$ are equivalent, which of course, as shown in Section 4.1, does not imply that their ε-entropies be of the same order.

It is of interest to consider some analytic consequences of the above results. One conjecture is that there is a relationship between the ε-entropy of a space and the speed of convergence of sequences of elements of the space. As far as we know, probabilists have not studied the speeds of convergence of corresponding sequences of distributions and characteristic functions, but have simply utilized Theorem 4.1 (i. e. $d(F_n, F) \to 0$ if and only if $f_n \to f$ uniformly in every bounded interval). In a subsequent paper we plan to investigate this problem using ε-entropy. In this connection we would like to refer to Dudley [7], who has studied the speed of convergence of sequences of empirical measures with the help of the so-called ε-δ-entropy, a notion which is closely related to ε-entropy.

REFERENCES

[1] R. Adler, A. Konheim and M. McAndrew, Topological entropy. *Trans. Amer. Math. Soc.*, t. **11**, 1965, p. 309-319.

[2] S. Chevet, p-ellipsoïdes de l^p; mesures cylindriques gaussiennes. *Les probabilités sur les structures algébriques*, p. 55-73. Centre National de la Recherche Scientifique, Paris, 1970.

[3] S. Chevet, A. Badrikian and P. Bernard, Application de la notion d'ε-entropie à la recherche de conditions d'existence de modifications continues de fonctions aléatoires. *Les probabilités sur les structures algébriques*, p. 33-41. Centre National de la Recherche Scientifique, Paris, 1970.

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

[5] E. I. Dinaburg, The relation between topological entropy and metric entropy (Russian). *Dokl. Acad. Nauk SSSR*, t. **190**, 1970, p. 19-22.

[6] R. M. Dudley, The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis*, t. **1**, 1967, p. 290-330.

[7] R. M. Dudley, Some uses of ε-entropy in probability theory. *Les probabilités sur les structures algébriques*, p. 113-122, Centre National de la Recherche Scientifique, Paris, 1970.

[8] R. M. Dudley, Small operators between Banach and Hilbert spaces. *Studia Math.*, t. **38**, 1970, p. 35-41.

[9] J. Dugundji, Topology. Allyn and Bacon, Boston, 1965.

[10] V. D. Erohin, Asymptotic theory of the ε-entropy of analytic functions (Russian). *Dokl. Akad. Nauk SSSR*, t. **120**, 1958, p. 949-952.

[11] B. V. GNEDENKO and A. N. KOLMOGOROV, *Limit Distributions for Sums of Independent Random Variables* (Translated from the Russian), Addison-Wesley, Reading, Mass., 1954.

[12] L. W. GOODWYN, Topological entropy bounds measure-theoretic entropy. *Proc. Amer. Math. Soc.*, t. **23**, 1969, p. 679-688.

[13] A. I. KHINCHIN, *Mathematical Foundations of Information Theory* (Translated from the Russian), Dover, New York, 1957.

[14] A. N. KOLMOGOROV, On the representation of continuous functions of several variables by superposition of continuous functions of a smaller number of variables (Russian). *Dokl. Akad. Nauk SSSR*, t. **108**, 1956, p. 179-182.

[15] A. N. KOLMOGOROV and V. M. TIHOMIROV, ε-entropy and ε-capacity of sets in function spaces (Translated from the Russian). *A. M. S. Translations*, série 2, t. **17**, 1961, p. 277-364.

[16] M. LOÈVE, *Probability Theory*, Third Ed., Van Nostrand, Princeton, N. J., 1963.

[17] G. G. LORENTZ, *Approximation of Functions*. Holt, Rinehart and Winston, New York, 1966.

[18] G. G. LORENTZ, Metric entropy and approximation. *Bull. Amer. Math. Soc.*, t. **72**, 1966, p. 903-937.

[19] E. LUKACS, *Characteristic Functions*, Griffin, London, 1960.

[20] E. C. POSNER and E. R. RODEMICH, Epsilon entropy and data compression. *Ann. Math. Statist.*, t. **42**, 1971, p. 2079-2125.

[21] R. T. PROSSER, The ε-entropy and ε-capacity of certain time-varying channels. *J. Math. Anal. Appl.*, t. **16**, 1966, p. 553-573.

[22] R. T. PROSSER, On the analysis and synthesis of certain abstract systems. *Bull. Amer. Math. Soc.*, t. **77**, 1971, p. 444-448.

[23] A. RÉNYI, *Probability Theory*. North-Holland, Amsterdam, 1970.

[24] J. RIORDAN, *Introduction to Combinatorial Analysis*. Wiley, New York, 1958.

[25] J. RIORDAN, *Combinatorial Identities*. Wiley, New York, 1968.

[26] V. STRASSEN and R. M. DUDLEY, The central limit theorem and ε-entropy. *Lecture Notes in Mathematics*, vol. **89**. Probability and Information Theory, p. 224-231. Springer-Verlag, Berlin, 1969.

[27] V. N. SUDAKOV, Gaussian measures, Cauchy measures and ε-entropy (Russian). *Dokl. Akad. Nauk SSSR*, t. **185**, 1969, p. 51-53.

[28] A. G. VITUŠKIN, On Hilbert's thirteenth problem (Russian). *Dokl. Akad. Nauk SSSR*, t. **95**, 1954, p. 701-704.

[29] A. G. VITUŠKIN, *Theory of Transmission and Processing of Information* (Translated from the Russian). Pergamon, New York, 1961.