AN ABSTRACT OF THE THESIS OF

Matthew Halfant for the degree Doctor of Philosophy

(Name of student) (Degree)

in Mathematics presented on June 19, 1974

(Major department) (Date)

Title: Some Results in the Ergodic Theory of Generalized

Expansions of Real Numbers

Abstract approved:

(Signature)

A. T. Lonseth

Let φ be a real-valued function defined on $I = [0,1]$.
We assume that (1) $\varphi(0) = 0$, $\varphi(1) = p$, an integer ≥ 2,
(2) $\varphi \in C^2[0,1]$, and (3) $\varphi'(x) > 0$ for $x \in I$. The
function φ may be used to define a generalization of the
ordinary base-p expansion, as follows. For $x \in [0,1)$, put
$a_1 = [\varphi(x)]$, $r_1 = \varphi(x) - a_1$. For $n \geq 1$, $a_{n+1} = [\varphi(r_n)]$,
$r_{n+1} = \varphi(r_n) - a_{n+1}$. The terms $\{a_n\}$ comprise the
expansion; the r_n are the remainders. In 1957, Alfréd
Rényi introduced an additional, technical hypothesis on φ,
thereby bringing ergodic theory into play and proving the
existence of a measurable function h such that, for almost
all $x \in I$, the sequence of remainders is distributed with
density h.

The present paper pursues two basic goals: one is to identify more clearly the class of functions satisfying Rényi's technical hypothesis; the other is to determine how smoothness conditions imposed on φ are reflected in h. Examples of results are: Rényi's hypothesis is shown to be met if $\varphi'(x) > 1$ for all $x \in I$; if further φ is analytic on I, so is h.

SOME RESULTS IN THE ERGODIC THEORY
OF GENERALIZED EXPANSIONS OF REAL NUMBERS

INTRODUCTION

This paper shall be concerned with a class of functions defined on the closed unit interval $I = [0,1]$, which are continuous, strictly increasing, and satisfy the boundary conditions $\varphi(0) = 0$, $\varphi(1) = p$, an integer ≥ 2. Such a function may be used to associate with each $x \in [0,1)$ an infinite sequence $\{a_n\}$ of integers in the following way. Take a_1 to be the greatest integer in $\varphi(x)$, and let $r_1 = \varphi(x) - a_1$ be the <u>remainder</u>. Now r_1 lies in $[0,1)$, and so we may iterate:

$$a_2 = [\varphi(r_1)], \qquad r_2 = \varphi(r_1) - a_2.$$

Generally, for $n \geq 1$,

$$a_{n+1} = [\varphi(r_n)], \qquad r_{n+1} = \varphi(r_n) - a_{n+1}.$$

If the reader will convince himself that the choice

$$(1) \qquad\qquad \varphi(x) = px$$

leads to the ordinary expansion of a number to base p, then he will begin to understand what Sôichi Kakeya had in mind when he introduced this "generalized expansion" scheme in 1924 [1]. Actually, Kakeya construed the function φ somewhat more generally: while required to be strictly monotonic, it could be either increasing or decreasing, and the upper limit of its range could be $+\infty$. The motivation for this was the fact that the special case

$$(2) \qquad \varphi(x) = 1/x$$

leads to the development of a number as a simple continued fraction; Kakeya wanted to unite this phenomenon with the base p expansion corresponding to (1) in a single generalization.

Now in order that the generalization be precise, it is necessary that we be able to _retrieve_ a number x from its associated expansion $\{a_1, a_2, \ldots\}$. Let f denote the function inverse to φ; then we have

$$x = f(a_1 + r_1) = f(a_1 + f(a_2 + r_2)) = \cdots = f(a_1 + f(a_2 + \cdots f(a_n + r_n) \cdots)).$$

We denote the final expression more simply by

$$x = [a_1, a_2, \cdots, a_n + r_n].$$

What is wanted is that

$$x = \lim_{n \to \infty} [a_1, a_2, \cdots, a_n].$$

To obtain a convenient formulation of this requirement, we use the notion of an _interval of rank n_, by which is meant a set of t satisfying an inequality of the form

$$[a_1, a_2, \cdots, a_n] \leq t < [a_1, a_2, \cdots, a_n + 1].$$

The elements belonging to this interval are characterized by having a_1 through a_n as the first n terms of their expansions. I is partitioned by p^n intervals of rank n, and the longest of these, I will say, has length $\triangle_n$. The condition

$$(3) \qquad \lim_{n \to \infty} \Delta_n = 0$$

guarantees that two different x's won't have identical expansions. For choose n so that Δ_n is smaller than the difference between these x's. Then the expansions cannot agree for the first n terms. Thus (3) states compactly that the expansions generated by φ are unique; a function φ for which this is true I shall call <u>generative</u>. Kakeya proved that <u>generativity</u> would obtain on the hypothesis that $|\varphi'(x)| > 1$ for almost every $x \in I$.

Kakeya's paper seems to have gone unnoticed by Western mathematicians. In 1944 B.H. Bissinger [2] generalized the continued fraction by considering decreasing functions other than (2); two years later, C.J. Everett [3] generalized (1) by considering just that class of functions I have described in my opening paragraph. By the way, Everett showed that there are generative functions with slopes numerically smaller than 1.

Kakeya's scheme opens the door to the question as to whether or not the integers that enter these generalized expansions make their appearance with certain well-defined frequencies. Thus, Borel showed early in this century that each of the digits 0,1,2,...,p-1 would occur with frequency 1/p in the base p expansion of almost every number x. An analogous result holds for the continued fraction, and has a colorful history.

Gauss stated in 1812 in a letter to Laplace (see Appendix III of Uspensky's book on probability [4]) that $\Phi_n(t)$, the distribution function of the nth remainder r_n of the continued fraction expansion (2), would converge as $n \to \infty$ to a limiting distribution function $\Phi(t)$, given explicitly by

$$\Phi(t) = \frac{\log(1+t)}{\log 2}.$$

But we do not possess his proof. It was only in the late 1920's that one was supplied: first, by R.O. Kuz'min [5] and then independently by Paul Lévy. Work on continued fractions was carried on into the thirties by Khinchin and by Lévy, and their books [6], [7] give readable accounts. Among many results, we find the assertion that the continued fraction expansion of almost every x will contain the integer k a proportion of times given by

$$\frac{1}{\log 2} \log\left(1 + \frac{1}{k(k+2)}\right) = \Phi\left(\frac{1}{k}\right) - \Phi\left(\frac{1}{k+1}\right).$$

(On Gauss' result, the right hand side represents the limiting probability, as $n \to \infty$, of the appearance of k as the nth term. However, we cannot pass immediately from this to the assertion just made because the terms $a_1, a_2, a_3, \ldots$ are not statistically independent.)

Alas, this and many other very exciting results depended on special features of the function 1/x, and they provide no useful clue as to how such results might be

extended to Kakeya's setting. Fortunately, there was another development taking place in the thirties: ergodic theory. The application of ergodic ideas to the continued fraction may be found in a lengthy paper of 1940 by Doeblin [8]; a more prominent display is given in the 1951 paper of Ryll-Nardzewski [9]. For a thrilling narrative account of this work, I recommend chapter 5 of Mark Kac's Carus monograph [10]; for further mathematical details, see section 4 of chapter 1 of the excellent book by Billingsley [11].

What is important here is not the relative ease with which hard-won results of Khinchin could now be recaptured by ergodic methods, but that such methods made possible the generalization from continued fractions to the domain of Kakeya's expansions. The decisive step was taken in 1957 by Alfréd Rényi [12]. Rényi worked with a class of functions satisfying conditions borrowed from Bissinger and Everett, and resembling somewhat Kakeya's, except that in the case of decreasing φ the derivative $|\varphi'(x)|$ (should it exist) was allowed to equal 1 on certain intervals, chosen so as not to disturb the generativity of the function φ. But in order to obtain results from ergodic theory, Rényi was forced to impose on φ a certain very technical condition (which I shall always refer to as Rényi's condition). Let t be a real variable, and define

$$H_n(x,t) = \frac{d}{dt}[a_1, a_2, \cdots, a_n + t]$$

where $x = [a_1, a_2, \ldots, a_n]$. Rényi's condition posits that there exist a constant $C \geq 1$, not depending either on n or on x, such that

$$(4) \qquad \frac{\sup_{t \in I} |H_n(x,t)|}{\inf_{t \in I} |H_n(x,t)|} \leq C.$$

On the strength of this hypothesis, Rényi proceeds to show

First: There exists a probability measure ν, equivalent to Lebesgue measure on I, invariant with respect to the transformation

$$T: \quad x \mapsto \varphi(x) - [\varphi(x)].$$

This means that for any (Lebesgue-) measurable set $E \subset I$,

$$\nu(T^{-1}E) = \nu(E).$$

Second: Let Φ be the associated invariant distribution,

$$\Phi(t) = \nu([0, t]).$$

Then the density $h(t) = \Phi'(t)$ (which I shall refer to as the invariant density) satisfies the inequality

$$\frac{1}{C} \leq h(t) \leq C, \qquad a.e. \ t \in I$$

the constant being the same as in (4).

Third: The transformation T is ergodic.

It now follows from the ergodic theorem that if F is any Lebesgue integrable function on I, then for almost every x,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} F(T^{k-1}x) = \int_I F \, d\nu = \int_0^1 F(t) \, h(t) \, dt.$$

Since $T^n x = r_n$, we obtain the frequency of the integer k by choosing, as F, the characteristic function of the interval $[f(k), f(k+1))$.

The focus of the present study is the subclass of those functions described at the outset which are <u>generative</u> and which in addition satisfy Rényi's condition. Such functions I will say belong to the class $\mathcal{R}$ of <u>Rényi functions</u>. I wish to show something of what the elements of $\mathcal{R}$ <u>look like</u>. The immediate obstacle is the apparent awkwardness of Rényi's condition.

Rényi himself verifies (4) in three cases: the base p expansion (1), which is trivial; the continued fraction case (2), for which special formulas are conveniently available; and in the case of the function

$$(5) \qquad \varphi(x) = (1+x)^m - 1, \qquad m \text{ an integer} \geq 2.$$

The latter case involves inversion of (5), and verification of (4) through a bit of manipulative virtuosity. But there is no indication as to how one might proceed with (5) replaced by

$$(6) \qquad \varphi(x) = (1+x)^m - 1 + \frac{\sin \pi x}{1000};$$

yet (6) does in fact give a function every bit as much an

element in $\mathcal{R}$ as does (5).

Part of the difficulty in characterizing the members of $\mathcal{R}$ -- indeed, of characterizing the generative functions -- stems from the level of generality assumed. Yet it has seemed to me perfectly natural to inquire: among the class of everyday functions that are <u>differentiable</u>, say, which are the members of $\mathcal{R}$? I have not hesitated to impose such conditions of differentiability as might be needed to render the question accessible to the methods of classical analysis.

To be specific, I intend to deal with what I shall call the class $\mathcal{A}$ of <u>admissible functions</u>. By definition, $\varphi \in \mathcal{A}$ if and only if these conditions are met:

$\mathcal{A}(1)$: $\varphi(0) = 0$, $\varphi(1) = p$.

$\mathcal{A}(2)$: φ has two continuous derivatives on $I = [0,1]$.

$\mathcal{A}(3)$: $\varphi'(x) > 0$ on I.

In requiring the continuity of the second derivative, the second of these conditions is stronger than it need be, and I have made it so principally to simplify the exposition. It would do as well to have the second derivative <u>piecewise</u> continuous. In fact, the interested reader will have no trouble in verifying that all the theorems remain correct with $\mathcal{A}(2)$ replaced by

$\mathcal{A}(2')$: φ' absolutely continuous, φ'' essentially

bounded, on I.

I will show quite explicitly that the requirement of boundedness is crucial.

In Part 1 I examine elements of $\mathcal{A}$ for membership in $\mathcal{R}$. Thus in Theorem 1 I show there is an auspicious subclass $\mathcal{A}_1 \subset \mathcal{A}$ whose elements are <u>always</u> in $\mathcal{R}$: $\mathcal{A}_1$ consists of the functions in $\mathcal{A}$ whose derivative remains > 1 on I. Observe that this dispatches (5) and (6) with equal ease. If the derivative can assume values ≤ 1, things are inherently more complicated, and the remainder of the section is devoted to a cursory exploration of the possibilities. It is shown, for example, that an admissible function φ may definitely fail to belong to $\mathcal{R}$ by virtue of the derivative φ' being equal to 1 at a single point. On the other hand, membership in $\mathcal{R}$ <u>need</u> not be spoiled by having φ' take on values ≤ 1. Thus, I extend Everett's nice result that there exist generative functions having arbitrarily small slope on an interval of length arbitrarily close to 1: I show that there are functions of the same description in $\mathcal{R}$.

Part 2 is concerned with the following sort of question: given that one imposes certain smoothness conditions on φ, what then can be concluded about the invariant density, h? Rényi shows that it is bounded away from 0 and from ∞, and that it is measurable. But is it <u>continuous</u>? Can we say when it must be differentiable?

As a means of approaching such issues, I begin by proving an analogue of the Gauss-Kuz'min theorem: if $\Phi_n(t)$ denotes the distribution of the nth remainder

$$r_n = T^n x,$$

then Φ_n converges uniformly to the invariant distribution Φ as $n \to \infty$. For the ideas of this proof I am indebted to Billingsley [11], who develops them in connection with the continued fraction; I have merely noted that the same ideas could serve in the more general setting of Rényi. Using this result, I am able to show that, at least for a certain well defined subset of the auspicious class $\mathcal{A}_1$, the differentiability of φ is indeed inherited by h so that, in particular, h is infinitely differentiable on I if φ is.

As a final result, I show that analytic elements of $\mathcal{A}_1$ give rise to analytic invariant densities, h.

For the convenience of both reader and writer, I post here the following list of definitions, notations, and elementary formulas upon which the text shall repeatedly draw.

The class $\mathcal{A}$ of admissible functions: $\varphi \in \mathcal{A}$ if and only if:

$\mathcal{A}(1)$: $\varphi(0) = 0$, $\varphi(1) = p$, an integer ≥ 2.

$\mathcal{A}(2)$: $\varphi \in C^2[0,1]$.

$$\mathcal{A}(3): \quad \varphi'(x) > 0 \quad \text{for } x \in I = [0,1].$$

<u>The auspicious subclass</u> $\mathcal{A}_1$: $\varphi \in \mathcal{A}_1$ if and only if:

$\varphi \in \mathcal{A}$ and $\varphi'(x) > 1$ for $x \in I$.

<u>Expansion notation</u>: f is the function inverse to φ. Let n be a positive integer, and let us momentarily regard λ_1, $\lambda_2, \ldots, \lambda_n$ as formal variables. We adopt the notation

$$f(\lambda_1 + f(\lambda_2 + \cdots + f(\lambda_n) \cdots)) = [\lambda_1, \lambda_2, \cdots, \lambda_n]$$

(except when n = 1 or 2, in which case I simply write things out the long way). The λ's must take values such that f does not receive an argument out of its domain. This will always be the case in the text, where typically λ_n will have some value in [0,p], while the other λ's are integers between 0 and p-1.

<u>Fundamental intervals</u>: For a given n, let $a_1, a_2, \ldots, a_n$ be given, where each a_i is one of the integers $0, 1, 2, \ldots, p-1$. The interval whose endpoints are

$$[a_1, a_2, \ldots, a_n] \quad \text{and} \quad [a_1, a_2, \ldots, a_n+1]$$

is called an <u>interval of rank n</u>. I is partitioned by p^n such intervals, the largest among them having length Δ_n. The collection of all intervals of rank $1, 2, 3, \ldots$ is called the set of <u>fundamental intervals</u>.

The condition of generativity: a function $\varphi \in \mathcal{A}$ is generative if and only if $\triangle_n \to 0$ as $n \to \infty$.

The transformation T: This is the p-to-1 map of I onto itself defined by

$$T: x \longmapsto \begin{cases} \varphi(x) - [\varphi(x)] & \text{for } 0 \le x < 1, \\ 1 & \text{for } x = 1. \end{cases}$$

The special assignment at x = 1 is not essential, but I adopt it in order to simplify the statement of certain theorems later on.

Rényi's condition: Let

$$H_n(x,t) = \frac{d}{dt}[a_1, a_2, \cdots, a_n + t]$$

where $x = [a_1, a_2, \ldots, a_n]$, each of the a's being an integer from 0 to p-1. Rényi's condition is that there exist a constant C such that, for any choice of n and x,

(7)
$$\frac{\sup_{t \in I} H_n(x,t)}{\inf_{t \in I} H_n(x,t)} \le C.$$

The class $\mathcal{R}$ of Rényi functions: $\varphi \in \mathcal{R}$ if and only if: $\varphi \in \mathcal{A}$, φ is generative, φ satisfies Rényi's condition.

The iterated distributions and densities: Let μ denote Lebesgue measure on I. For n = 0,1,2,... I define the <u>nth iterated distribution</u> Φ_n by

$$(8) \qquad \Phi_n(t) = \mu(\{x \in I: r_n = T^n x \leq t\}) = \mu(T^{-n}[0,t]).$$

The <u>iterated densities</u> S_n are the derivatives

$$S_n(t) = \Phi_n'(t).$$

The Φ_n may be written out explicitly as

$$\Phi_n(t) = \sum_{a_1=0}^{p-1} \sum_{a_2=0}^{p-1} \cdots \sum_{a_n=0}^{p-1} \left\{ [a_1, a_2, \cdots, a_n + t] - [a_1, a_2, \cdots, a_n] \right\}$$

or recursively, by

$$(9) \qquad \begin{aligned} \Phi_0(t) &= t; \\ \Phi_{n+1}(t) &= \sum_{k=0}^{p-1} \left\{ \Phi_n(f(k+t)) - \Phi_n(f(k)) \right\} \quad \text{for } n \geq 0. \end{aligned}$$

The S_n may be described recursively by

$$(10) \qquad \begin{aligned} S_0(t) &= 1; \\ S_{n+1}(t) &= \sum_{k=0}^{p-1} S_n(f(k+t)) f'(k+t) \quad \text{for } n \geq 0 \end{aligned}$$

or explicitly by

$$(11) \qquad S_n(t) = \sum_x H_n(x, t)$$

where the summation is taken over all $x = [a_1, a_2, \ldots a_n]$ as each a_i ranges independently from 0 to p-1.

Invariant measure, distribution, and density: The

invariant measure ν is a probability measure on I, whose existence is guaranteed for a function $\varphi \in \mathcal{R}$, such that for an arbitrary measurable set $E \subset I$,

$$\nu(T^{-1}E) = \nu(E).$$

The invariant distribution Φ is defined by

$$\Phi(t) = \nu([o,t]) \qquad t \in I,$$

and the invariant density h is defined almost everywhere on I by

$$h(t) = \Phi'(t).$$

Rényi has shown:

(12) $$\frac{1}{C} \leq h(t) \leq C, \qquad a.e.\ t \in I.$$

PART 1

How do we suppose that Archimedes would have calculated the decimal expansion of $\sqrt{2}$? That is — assuming he knew analytic geometry. The answer is obvious: first he would connect the origin to the point (1,10) with a straight line. Then he would place a compass point at (-1,0) and adjust the pencil to meet (0,1). The arc brought down from this point would, of course, hit the x-axis at $\sqrt{2}$ - 1. Now the fun begins. A vertical erected at this intercept crosses the line $y = 10x$ between $y = 4$ and $y = 5$; Archimedes calls out "IV" to his scribe, and sets a pair of dividers to record the remainder. Then placing one point of the instrument at the origin, he marks off the remainder on the x-axis, and proceeds to the next digit. "For what has been done once, can always be repeated." This is ideal geometry: lines with no thickness, settings with no error to grow tenfold at each repetition. We may imagine our friend Archimedes continuing indefinitely to call out the digits of $\sqrt{2}$: 1.41421356... .

Now let me thicken the plot. Enter a mischievous spirit: in an instant, the line segment joining the origin and (1,10) disappears, and is replaced by an arc of an extremely large circle — to be specific, a circle of radius 10^{100}, with its center below the horizontal axis, in

the fourth quadrant. Now if Archimedes were to sight along this "line" by making his eye colinear with the origin and (1,10), then he would certainly detect the upward bulge that betrays the work of a rascal. But since he has no reason to do this, I shall suppose that the change goes completely unnoticed, Archimedes continuing to read off his digits by use of the arc that stands now in place of the line $y = 10x$.

What happens? As the reading of the digits continues, there will be at first no discrepancy to give evidence of the fraud. Then suddenly the digits will go "bad". But will they be obviously bad? Could one tell by looking at them that they were bad? The answer is "Yes", provided we suppose that $\sqrt{2}$ is a "normal" number -- in which all the digits occur with equal frequency -- and that the scribe is Émile Borel. For Borel would be bound to notice, over the aeons, a slight but persistent preponderance of 9's over 0's. This is because the arc is concave down; were it concave up, with the center in the second quadrant, it would be the smaller digits that would be favored in the expansion.

Very well, now what is to be made of all this? We start with a line segment $y = 10x$ (or $y = px$, more generally) with x going from 0 to 1. Now, while holding the endpoints fixed, we effect some smooth distortion of the segment, giving a curve $y = \varphi(x)$. We suspect that

small distortions will have small effects upon the behaviour of the expansions, and it is even suggested to our intuition that quite noticeable distortions (corresponding to decidedly nonlinear $\varphi(x)$) may not impair the qualitative character of the resulting expansions. By "qualitative character" I mean to embrace two things: (a) the uniqueness of expansions, for representing numbers, and (b) the statistical regularity of the various digits. Thus, while we know that the digits will not continue to be equally likely, we feel that new "distorted" probabilities must have evolved from the former, equal ones.

The mathematical question at issue, then, is to see how freely our $\varphi(x)$ can be chosen, without our losing the qualities expressed by (a) and (b). As regards the former of these qualities -- that which I call generativity -- a very satisfying answer has been given by Kakeya. He instructs us to look at the <u>slope</u>, for the lines y = px tend to be steep, all slopes ≥ 2. Kakeya says: <u>Any smooth deformation will not destroy generativity provided the slope is always kept greater than 1.</u> (The trouble that can occur with slopes less than 1 is real; the reader may wish to explore this for himself by attempting to geometrically construct the fundamental intervals for a function like $\varphi(x) = 2x^2$.)

But what about question (b)? That is the point of the

following theorem: to show that Kakeya's answer again
applies.

Theorem 1: $\mathcal{A}_1 \subset \mathcal{R}$.

Proof: Since $\varphi'(x)$ is continuous on a closed interval, it
attains therein a maximum and a minimum value; in
particular, the minimum value is strictly greater than 1.
Equivalently, we may say of f'(x) that

$$0 < \alpha \leq f'(x) \leq \beta < 1 \qquad \text{for all } x \in [0,p].$$

Let me use this now to show that

$$\Delta_n \leq \beta^n$$

for all n. First, a typical interval of rank 1 has length

$$f(a_1 + 1) - f(a_1) = f'(\theta)$$

for some $\theta \in (a_1, a_1+1)$, by the mean value theorem. Thus Δ_1
$\leq \beta$. Proceeding inductively, we assume $\Delta_k \leq \beta^k$ for
$1 \leq k < n$. An interval of rank n will have length given by

$$f(a_1 + [a_2, \cdots, a_n+1]) - f(a_1 + [a_2, \cdots, a_n])$$

$$= f'(\theta)\left([a_2, \cdots, a_n+1] - [a_2, \cdots, a_n]\right) \leq \beta \Delta_{n-1} \leq \beta^n,$$

and the result is shown.

On the strength of our hypotheses, we see that

$$H_n(x,t) = f'(a_1 + [a_2, \cdots, a_n + t]) \, f'(a_2 + [a_3, \cdots, a_n + t]) \cdots f'(a_n + t)$$

will actually assume its sup and inf at two points t_1 and t_2 belonging to I. Now the number

$$C_n(x) = \frac{H_n(x, t_1)}{H_n(x, t_2)}$$

can be expressed as the product of n fractions, each having the form

(13)
$$\frac{f'(b + [b_1, b_2, \cdots, b_k + t_1])}{f'(b + [b_1, b_2, \cdots, b_k + t_2])} \quad , \qquad 1 \le k < n.$$

The case where k = 0 corresponds to the factor

$$\frac{f'(a_n + t_1)}{f'(a_n + t_2)} \le \frac{\beta}{\alpha} \, .$$

In the general term displayed in (13), the arguments of f' differ by a number d_k not larger in magnitude than $\triangle_k$. By the mean value theorem, we may rewrite this fraction as

(14)
$$\frac{f'(b + [b_1, \cdots, b_k + t_2]) + f''(\theta) \, d_k}{f'(b + [b_1, \cdots, b_k + t_2])} \, .$$

f" is continuous on [0,p], and so there exists an M such that $|f''(x)| \le M$ for all $x \in [0,p]$. Thus (14) can be rewritten as

(15) $\qquad 1 + \varepsilon_k$, with $|\varepsilon_k| \le \dfrac{M}{\alpha} \Delta_k \le \dfrac{M}{\alpha} \beta^k$.

Putting all this together, we have

$$C_n(x) \le \frac{\beta}{\alpha} \prod_{k=1}^{n-1} \left(1 + \frac{M}{\alpha} \beta^k \right).$$

The product on the right hand side converges as $n \to \infty$, by virtue of the convergence of the geometric series

$$\sum_{k=1}^{\infty} \beta^k.$$

Therefore Rényi's condition holds with the constant

$$C = \frac{\beta}{\alpha} \prod_{k=1}^{\infty} \left(1 + \frac{M}{\alpha} \beta^k \right).$$

Now we know that we are safe as long as the slope stays > 1. But having gotten this far, it becomes compelling to ask if the "boundary" of allowable slope is real or imagined. We may argue that it _must_ be real, since it is so for the more basic question of generativity -- still we want to know just what it is that goes wrong when slopes get ≤ 1. To that end, consider as a simple example the function

(16) $\qquad\qquad \varphi(x) = x + x^2$.

Since $\varphi'(x) = 1 + 2x$, we see that the slope is bigger than 1 _except_ when $x = 0$; here the function bravely sticks its toe in the water to see what happens. I would recommend that the interested reader try playing with this on a desk calculator; after studying the pattern of remainders for a while, certain things begin to appear. For one, the succession of remainders is not always haphazard, as might be expected. We note that any occurrence of a _small_ remainder will initiate an entire "run" of small remainders, each slightly larger than the last. This regularity is most pronounced when we encounter an r_n which is quite close to zero; for then we have $r_{n+1} = Tr_n = r_n + r_n^2$, where it is clear that the increase is hardly perceptible. From this point on, the successive remainders will gradually build themselves back up by ever increasing increments -- the process at this stage is certainly not "random". The randomness will reappear when the greatest integer function comes back into play, and this will happen when the sequence of remainders has sufficiently escaped the vicinity of the origin to first enter the interval $[f(1),1)$. As the increments have all the while been increasing in size, the exact place of this first landing in $[f(1),1)$ will be completely unpredictable; that is, we _observe_, at this occasion, the generation of a _new_ remainder, "truly" random in I. After this, the successive remainders will continue to bounce "more or less" freely

over I, until again one happens to land near the origin.

Let me contrast this now with the kind of behaviour to be expected from a Rényi function. Suppose that (16) is a function in $\mathcal{R}$, and let h be the invariant density. The ergodic theorem says that for almost every x, the sequence or remainders

$$(17) \qquad\qquad r_1, \; r_2, \; r_3, \; \cdots$$

is <u>distributed</u> on I with density h. What is more, we know from the inequality (12) that h is (essentially) bounded away from 0 and from ∞, and this means that the terms of (17) cannot start piling up excessively in any one section of I. Yet we have just seen how the remainders for (16) are <u>detained</u> in the vicinity of the origin; it is, in fact, on this account that the function (16) is excluded from $\mathcal{R}$. I will prove this rigorously later on, but for now let me continue in this informal vein and offer a rough argument which I myself find more convincing than the proof.

Let us visualize the sequence (17) through the agency of a "particle" which appears in succession at the appropriate locations in I. We notice that a single visit to a neighborhood lying just to the right of f(1) is followed by a sojourn in the vicinity of the origin; this suggests comparing h at f(1) with h at 0. To initiate the precise (as opposed to qualitative) argument, we agree to

measure the degree to which a particle finds itself trapped, by the number of steps it takes to double its distance from the origin: in precise terms, I define $N(\varepsilon)$ to be the smallest positive integer k for which $T^k\varepsilon \geq 2\varepsilon$. I use ε here to suggest that we are interested in what happens for starting positions that are <u>close</u> to the origin. Thus ε will be small, and therefore we have

$$T\varepsilon = \varepsilon + \varepsilon^2, \qquad T(2\varepsilon) = 2\varepsilon + 4\varepsilon^2.$$

If x should lie between ε and 2ε, then the increment it acquires from action by T will lie between ε^2 and $4\varepsilon^2$; this gives us the estimate, valid for ε sufficiently small,

$$(18) \qquad \frac{1}{4\varepsilon} < N(\varepsilon) < \frac{1}{\varepsilon}.$$

Now let us consider the two minute intervals

$$J_0 = [0, 2\varepsilon], \qquad J_1 = [f(1), f(1+\varepsilon)].$$

(The significance of the second of these will be clear in a moment.) For almost every $x \in I$, the sequence of remainders $r_n = T^n x$ must dip into these two intervals so that the ratio of visits to J_0 by those to J_1 must approach, as a limit,

$$(19) \qquad \frac{\int_{J_0} h \, d\mu}{\int_{J_1} h \, d\mu};$$

and if we do not worry about the possibility of h being discontinuous, we see that for small ε, (19) is sensibly the same as

$$\frac{h(0)\,\mu(J_o)}{h(f(1))\,\mu(J_1)} \;\approx\; \frac{2\,h(0)}{h(f(1))\,f'(1)}\;.$$

In any case, it is clear that the expression (19) remains finite as $\varepsilon \to 0$. However, the verdict of (18) won't allow this. For <u>whenever</u> the particle should enter J_1, it is straightaway whisked by T to $[0, \varepsilon]$, and so must remain in J_0 for at <u>least</u> $N(\varepsilon)$ steps. Thus, according to (18), the degree to which J_0 is favored over J_1 must increase without bound as $\varepsilon \to 0$, and here the contradiction is revealed.

By the way, this argument does not address itself to the question of the <u>digits</u> in the expansions generated by (16). Whether the 0's and 1's appear with statistical regularity is something that I do not know, though I suspect that they do. I should point out that questions which arose in consideration of the <u>digits</u> have served principally to focus attention on the succession of <u>remainders</u>. The center of gravity of the problem has shifted now, with the transformation T (carrying each remainder into the next) and the density h being the objects of mathematical interest. What the heuristic argument given above does show is that the density h belonging to (16) must either become infinite at 0, or

vanish at f(1). Although this does not preclude regularity for the digits themselves, it does represent a genuine departure from the behaviour that characteristically attends the use of Rényi functions. For this class $\mathcal{R}$ of functions is intended to bear a kind of topological kinship to the linear variety $\varphi(x) = px$ for which the densities h are always uniform; the h resulting from a Rényi function must be "quasi-uniform" in the sense implied by the inequality (12).

In what follows, I would like to explore the pathology of admissible functions which do not belong to the auspicious class $\mathcal{A}_1$. In the first place, I shall consider functions φ whose derivative assumes a minimum value of 1 at a finite number of points; ultimately, the question of smaller slopes will be taken up.

It shall be necessary first to forge some tools, the principal one being already implicit in the proof of Theorem 1. By examining the final stages of the argument -- in particular, the inequality in (15) -- we see that a much more general result may be inferred; namely,

<u>Theorem 2</u>: $\quad \varphi \in \mathcal{A}, \quad \sum_{n=1}^{\infty} \Delta_n < \infty \quad \Rightarrow \quad \varphi \in \mathcal{R}.$

The condition

$$(20) \qquad \sum_{n=1}^{\infty} \Delta_n \; < \; \infty$$

is quite interesting, for it embraces the requirement (3) of generativity, but goes on to say what further is required for membership in $\mathcal{R}$. The function φ in (16) is generative according to Kakeya, from which we infer that while the Δ_n must approach 0, they somehow fail to do so sufficiently rapidly for the series in (20) to converge.

My fascination with (20) lead me quite early to a fairly instructive error, which I proceed now to describe. When I first began my study of these problems, I made use of an electronic calculator to experiment with various functions φ. It was in this way that I observed the behaviour of $\varphi(x) = x + x^2$ already discussed. At the time, however, I was not quite sure whether this function belonged to $\mathcal{R}$ or not. I noticed that the delays encountered near the origin were much more severe with the function

$$(21) \qquad \varphi(x) \; = \; x + x^3$$

than with the φ of (16), and I formed the idea that the order of contact of φ with the line $y = x$ must play a decisive role. To take an exaggerated case, consider the function

$$(22) \qquad \varphi(x) = x + e^{1 - \frac{1}{x^2}}.$$

If one should here encounter a remainder as small as .1, then the analogue of (18) for this function shows that more than $e^{24}/10 \approx 2\frac{1}{2}$ billion iterations would be required to reach .2; and this is more of a life-sentence than a delay. Thus I was prepared to believe that (22) did not belong to $\mathcal{R}$, while (16) and (21) ... maybe.

It was about this time that I found a way to relate the order of contact between $\varphi(x)$ and $y = x$ with the way in which the $\triangle_n$ approached 0. Consider an admissible function $\varphi_0(x)$ whose derivative is equal to 1 at the single point $x = 0$, and suppose that $\varphi_0''(x) > 0$ for $x > 0$. Then it can be shown -- I do so in the appendix -- that (20) will be met provided the integral

$$(23) \qquad \int_\varepsilon^1 \frac{x}{\varphi_0(x) - x} \, dx$$

remains finite as $\varepsilon \to 0$ from above. Here again I found the suggestion that a function φ_0 of the type considered might indeed have a chance to be in $\mathcal{R}$ if only the curve $y = \varphi_0(x)$ could "pull away" sufficiently fast from the line $y = x$. For example, the integral in (23) remains finite if

$$(24) \qquad \varphi_0(x) = x + x^s$$

where s is a fixed constant lying strictly between 1 and 2. In this event $\sum_{n=1}^{\infty} \Delta_n$ does converge, and for a while I believed that the function in (24) was a bona fide member of $\mathcal{R}$. But arguments already advanced show that this cannot be so; where lies the error? Simply here: in taking s in (24) to be less than 2, the second derivative is caused to explode at x = 0 (although I must say, it did so very quietly). Thus the function is not admissible, and Theorem 2 cannot be applied. For all that, we do make one modest gain: we see that the boundedness of the second derivative is indeed a crucial requirement.

While the attack on the series appearing in (20) has yielded very little, it turns out that a direct and elementary estimate of the individual terms Δ_n proves fruitful. The following simple result is the natural companion of Theorem 2.

<u>Product lemma</u>: For each n, there exists a $\xi \in (0,1)$ such that

$$(25) \qquad \frac{1}{\Delta_n} = \prod_{k=1}^{n} \varphi'(T^{k-1} \xi).$$

<u>Proof</u>: For a certain choice of the a's, we have

$$\Delta_n = [a_1, a_2, \cdots, a_n + 1] - [a_1, a_2, \cdots, a_n]$$

$$= \frac{d}{dt} [a_1, a_2, \cdots, a_n + t] \Big|_{t = t^*}$$

for some $t^* \in (0,1)$, by the mean value theorem. Thus

$$\Delta_n = f'(a_1 + [a_2, \cdots, a_n + t^*]) \, f'(a_2 + [a_3, \cdots, a_n + t^*]) \cdots f'(a_n + t^*)$$

$$= \prod_{k=1}^{n} f'(\varphi(T^{k-1}\xi)) = \prod_{k=1}^{n} \frac{1}{\varphi'(T^{k-1}\xi)}$$

where $\xi = [a_1, a_2, \ldots, a_n + t^*]$.

Observe how consideration of (20) and (25) together leads to an easy rederivation of Theorem 1. But of far greater interest is the way in which (25) sheds light on the question of what happens when φ' takes values ≤ 1. What is to me quite remarkable is that formula (25) directs us to a consideration of sequences of the form

$$(26) \qquad\qquad \xi, \ T\xi, \ T^2\xi, \ \cdots \ .$$

We refer to (26) as the <u>orbit</u> of ξ under T. This is nothing new; the sequence of remainders (17) is an orbit under T, and it is on this very basis that ergodic theory can be applied. But formula (25) does not come from ergodic theory; it is purely classical. Still it speaks to us of orbits, saying, "If you want to make Δ_n small,

see to it that the orbit of ξ spends most of its time
where the slope φ' is bigger than 1." We must take this
to be a statement about orbits in general, since the ξ in
(25) is different for every n, and never explicitly known
anyway.

Let us see now what bearing this has on the question
of an admissible function φ whose derivative takes a
minimum value of 1 at a single point, x_o. In the case of
the functions (16), (21), and (22), we have $x_o = 0$ -- and
this is always an invariant point under T. That means that
we cannot get favorable estimates from (25), and although
we could neither conclude from this that φ does not belong
to $\mathcal{R}$, it is sufficient to point a suggestive finger at the
source of the trouble. For suppose, on the other hand,
that x_o were not an invariant point under T; that is, Tx_o
$\neq x_o$. Then although the orbit of ξ could land right on
top of x_o, it couldn't just remain there. If we try now to
get a useable estimate from (25), we see that this is not
quite enough; to bound the factors of (25) away from 1, we
must further concern ourselves with the possibility of an
orbit passing arbitrarily close to x_o. To get a hold on
this, we suppose that we can cover x_o with a small open
interval G which is disjoint from its image under T. Thus:
$x \in G \implies Tx \notin G$. Outside of G we must have φ' bounded
below by some number β strictly larger than 1. Then we may
be sure that at least every second factor in (25) is $\geq \beta$,

with the result that

$$\frac{1}{\triangle_n} \geq \beta^{\left[\frac{n}{2}\right]}.$$

From this we see that (20) is satisfied, so that $\varphi \in \mathcal{R}$ by Theorem 2.

Now we understand that the examples (16), (21), and (22) are somewhat fortuitous: they fail to be in $\mathcal{R}$ not because they all have slopes of 1 at a given point, but rather because this point happens to be invariant under T. I have said that $x_0 = 0$ is always invariant; the same applies to $x_0 = 1$, for <u>however</u> we define T at 1, we can never get a neighborhood of 1 to be disjoint from its image under T. Thus the definition of T on page 12 is phrased so as to make the invariance of 1 explicit, and the advantage thereby gained is that a non-invariant point may always be embedded in a "non-invariant" open set, as required in the argument above.

Let me present that argument now as a formal principle.

<u>Theorem 3</u>: $\varphi \in \mathcal{A}$, $\varphi'(x) \geq 1$ on I. Let U = $\{ x : \varphi'(x) = 1 \}$ and let G be an open set in I that covers U. Suppose there exists an N such that for every $x \in G$, $T^k x \in G^c$ for some $k \leq N$. Then $\varphi \in \mathcal{R}$.

<u>Proof</u>: Let $\beta = \inf \varphi'(x)$ over all $x \in G^c$; then $\beta \geq 1$.

Suppose $\beta = 1$. By the continuity of φ', and the fact that G^c is closed, there must be a point $x_0 \in G^c$ at which the infimum $\beta = 1$ is assumed. But then $x_0 \in U$, which is a contradiction. Thus $\beta > 1$. Now the product lemma can be invoked as before: since no point can remain in G under repeated operation by T for more than N steps, we have

$$\frac{1}{\Delta_n} \geq \beta^{\left[\frac{n}{N+1}\right]}$$

and the conclusion follows.

Let me now suppose that the set U contains at most a finite number of points. Then the following theorem, which gives a necessary and sufficient condition for membership in $\mathcal{R}$, should seem perfectly natural in light of the previous discussions.

Theorem 4: $\varphi \in \mathcal{A}$, $\varphi'(x) \geq 1$ on I. Suppose the set U on which $\varphi'(x) = 1$ consists of finitely many points. Then $\varphi \in \mathcal{R}$ if, and only if, U contains no orbit under T.

Proof: The "if" part of this theorem can be gotten from Theorem 3: it is only a question of constructing the set G. Suppose U to consist of the points

(27) $\qquad\qquad x_1, x_2, \ldots, x_m$

and suppose further that the points

(28) $$Tx_1, \; Tx_2, \; \ldots \; , \; Tx_m$$

are all distinct from those of U. Put

$$\varepsilon \; = \; \min_{i,j} \; \left| \, x_i - Tx_j \, \right|$$

and let $B = \sup_{t \in I} \varphi'(t)$. With x_i as midpoint, construct an
open interval G_i of length $\varepsilon/2B$. The image of G_i under T
cannot have length exceeding $\varepsilon/2$, from which we see that
the set

$$G \; = \; \bigcup_{i=1}^{m} \; G_i$$

is necessarily disjoint from its image under T.

Of course it could happen that certain terms in (28)
equal certain other terms in (27). For example, suppose

$$x_2 = Tx_1 \quad \text{and} \quad x_3 = Tx_2$$

but that Tx_3 does not belong to U (we have to get out of U
in m or fewer steps, since otherwise U contains an orbit).
For simplicity, suppose also that no other such
duplications occur. Let ε be the smallest distance
between the <u>distinct</u> points among (27) and (28) combined.
Center about x_1 an open interval G_1 of length $\varepsilon/2B^3$, and
let G_2 and G_3 be the successive images of G_1 under T. The
points $x_4, \; \ldots \; , \; x_m$ get covered by intervals of length

$\varepsilon/2B$; then we can say that any point in $G = \bigcup_{i=1}^{m} G_i$ must escape G under one, two, or three applications of T. Thus Theorem 3 can be applied with N = 3; and we see that the argument is general.

This brings us to the "only if" part of the theorem. Recall that in our earlier discussion of how a function φ might fail to belong to $\mathcal{R}$, we found that there were circumstances under which the invariant density h could not remain simultaneously bounded away from 0 and from ∞ -- a violation of the inequality (12). The idea employed in what follows is very similar, but it is applied, not to h, but rather to the iterated densities S_n. These latter are closely related to h; on the basis of a little experimenting it appears that the S_n must converge to h as $n \longrightarrow \infty$, and in Part 2 I shall prove this to be the case under appropriate conditions. Thus there should be no great surprise occasioned by the following lemma, in which it is shown that the iterated densities themselves are constrained by the same inequality that bounds the invariant density, h.

Iterated densities lemma: If $\varphi \in \mathcal{R}$, then for every $n \geq 0$,

$$\frac{1}{C} \leq S_n(t) \leq C \qquad \text{for all } t \in I.$$

<u>Proof</u>: Combining (7) and (11) gives

$$\sup_{t \in I} S_n(t) \leq C \inf_{t \in I} S_n(t).$$

But since $\int_0^1 S_n(t)dt = 1$, we must have $\inf_{t \in I} S_n(t) \leq 1$, so that

$$S_n(t) \leq C \qquad \text{for } t \in I.$$

Likewise, $\sup_{t \in I} S_n(t) \geq 1$, so that

$$S_n(t) \geq \frac{1}{C} \qquad \text{for } t \in I,$$

and the lemma is proved.

This result -- which to me is the intuitive content of Rényi's condition -- will enable us now to complete the proof of Theorem 4. Begin with the case in which U contains a single invariant point x_0 (that is, an orbit of cycle length 1). Notice that from $Tx_0 = x_0$ we may write $x_0 = f(j + x_0)$ where j is an integer in the range 0 to p-1: for if $x_0 < 1$, we take $j = [\varphi(x_0)]$; we take $j = p-1$ if $x_0 = 1$. Substituting $t = x_0$ in the recurrence formula

$$(10) \qquad S_{n+1}(t) = \sum_{k=0}^{p-1} S_n(f(k+t)) f'(k+t)$$

we obtain

$$(29) \qquad S_{n+1}(x_0) - S_n(x_0) = \sum_{k \neq j} S_n(f(k+x_0)) f'(k+x_0)$$

by virtue of the fact that $f'(j + x_0) = 1/\varphi'(x_0) = 1$.

Thus $S_n(x_0)$ increases with n. If φ is to be in $\mathcal{R}$, then the iterated densities lemma implies that the limit of $S_n(x_0)$ as $n \rightarrow \infty$ must exist and be $\leq C$; but then on summing (29) from $n = 0$ to ∞ we infer that for each $k \neq j$ we must have

$$\lim_{n \rightarrow \infty} S_n(f(k+x_0)) = 0.$$

Thus S_n cannot be bounded away both from 0 and from ∞, whence we can only conclude that $\varphi \notin \mathcal{R}$.

Similarly, suppose that U contains an orbit of length 2: e.g., $U = \{x_0, x_1\}$ where $x_1 = Tx_0$ and $x_0 = Tx_1$. Then we have, say, $x_0 = f(i + x_1)$ and $x_1 = f(j + x_0)$; in place of (29) we have the two equalities

$$S_{n+1}(x_0) - S_n(x_1) = \sum_{k \neq j} S_n(f(k+x_0)) f'(k+x_0),$$

$$S_{n+1}(x_1) - S_n(x_0) = \sum_{k \neq i} S_n(f(k+x_1)) f'(k+x_1).$$

We add these equations. Then if we want the terms

$$S_n(x_0) + S_n(x_1)$$

to remain bounded as $n \rightarrow \infty$, we are again forced to conclude that S_n must at certain points approach 0 as $n \rightarrow \infty$. Thus the truth of Theorem 4 is clear.

For the remainder of this section, I would like to deal with the unrestricted case in which the slope φ' may take on arbitrary positive values. Let V denote the set of those $x \in I$ for which $\varphi'(x) \leq 1$. We can see at once how much richer the problem has become if we consider trying to establish here an analogue of Theorem 3. There it was sufficient that an orbit should be able to "get away" from U every now and again, but clearly no such counting argument will work in the present context. Contemplation once more of the product lemma (25) shows that a delicate balance must be established between the orbit's conduct in and out of V. The following line of thought seems almost inescapable: Suppose $\varphi \in \mathcal{R}$; then according to the ergodic theorem, for almost every $x \in I$ we have

$$(30) \qquad \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \log \varphi'(T^{k-1} x) = \delta$$

where

$$\delta = \int_{0}^{1} \left\{ \log \varphi'(t) \right\} h(t)\, dt.$$

From (25) we have

$$(31) \qquad -\frac{1}{n} \log \Delta_n = \frac{1}{n} \sum_{k=1}^{n} \log \varphi'(T^{k-1} \xi)$$

and comparison with (30) now suggests the asymptotic relation

$$(32) \qquad \Delta_n \sim e^{-\delta n} \qquad \text{as } n \to \infty.$$

But there are real problems, the least of these being that
the conclusion (32) is a little too strong even if the
right hand side of (31) $\underline{does}$ approach the limit δ. More
serious difficulties are that the ξ in (31) is really ξ_n;
and that possibly none of these ξ_n is contained among the
x for which (30) holds. Still, the fact that we are lead
to write down (32) is enough to suggest that the condition

$$(33) \qquad\qquad \delta > 0$$

must have $\underline{some}$ significance; but my several attempts to
establish either its necessity or sufficiency for
membership in $\mathcal{R}$ have been entirely unavailing; and I
expect that the "boundary" condition $\delta = 0$ poses a far
subtler problem than (33).

What I $\underline{have}$ been able to show is offered in the
following two theorems. The first of these gives a
necessary condition for φ to be in $\mathcal{R}$: it is $\underline{necessary}$
that V contain no orbit under T. This constitutes a mild
extension of the "only if" part of Theorem 4; there the
orbit contained in U was necessarily cyclic, and the fact
that it needn't be so here means that the method of proof
must be slightly modified. The second theorem gives a
$\underline{sufficient}$ condition; it amounts to a generalization of
Theorem 1.

<u>Theorem 5</u>: Let $V = \{x : \varphi'(x) \leq 1\}$. Then if V contains an orbit under T, $\varphi \notin \mathcal{R}$.

<u>Proof</u>: We are given an $x_0 \in V$ with $x_n = T^n x_0 \in V$ for all $n \geq 0$. We may write $x_{n+1} = Tx_n = \varphi(x_n) - i_n$, so $x_n = f(i_n + x_n)$. Substitution of $t = x_{n+1}$ in (10) yields

$$(34) \quad S_{n+1}(x_{n+1}) = S_n(x_n) f'(\varphi(x_n)) + \sum_{k \neq i_n} S_n(f(k+x_{n+1})) f'(k+x_{n+1}).$$

Since $x_n \in V$, $f'(\varphi(x_n)) = 1/\varphi'(x_n) \geq 1$. Thus if we subtract $S_n(x_n)$ from both sides of (34), all terms on the right will remain non-negative. The summability of

$$S_{n+1}(x_{n+1}) - S_n(x_n)$$

is, as before, incompatible with the condition $S_n(t) \geq \frac{1}{C}$; hence the theorem is proved.

<u>Theorem 6</u>: For $n \geq 1$, let $\mathcal{A}_n$ denote the subclass of $\mathcal{A}$ whose elements φ satisfy

$$\inf_{\xi \in I} \prod_{k=1}^{n} \varphi'(T^{k-1} \xi) > 1.$$

Then each $\mathcal{A}_n \subset \mathcal{R}$.

<u>Proof</u>: (I should point out first that $\mathcal{A}_1$ retains its original meaning.) Let

$$\inf_{\xi \in I} \prod_{k=1}^{n} \varphi'(T^{k-1}\xi) = \beta > 1$$

and

$$\inf_{\substack{\xi \in I \\ m < n}} \prod_{k=1}^{m} \varphi'(T^{k-1}\xi) = \alpha.$$

Then

$$\frac{1}{\Delta_N} \geq \alpha \beta^{\left[\frac{N}{n}\right]}$$

so that $\sum \Delta_N < \infty$ and Theorem 2 applies.

A question brought to mind by the last result is this: if φ belongs to $\mathcal{R}$, does it necessarily belong to <u>some</u> $\mathcal{A}_n$? If it does, then we can easily show that it belongs as well to all the sets $\mathcal{A}_N$, $\mathcal{A}_{N+1}$, $\mathcal{A}_{N+2}$, ... for some N $\geq$ n. Thus the question may equivalently be worded: does $\varphi \in \mathcal{R}$ imply that φ belongs to all the $\mathcal{A}$'s from a certain point onward? The conjecture that it does may be written thus:

(?) $$\mathcal{R} = \bigcup_{n=1}^{\infty} \mathcal{A}_n.$$

It is not hard to show that the truth of (?) would imply that (33) <u>is</u> in fact a necessary condition for membership in $\mathcal{R}$; I do not think it would be so easy to prove the

implication going the other way around. Yet my suspicion that (?) is true derives largely from a feeling that it ought to follow from the (unproved!) necessity of (33).

But enough of this. Let me close this section on a positive note, with the following application of Theorem 6.

Theorem 7: There exist functions in $\mathcal{R}$ whose derivative remains arbitrarily small on an interval of length arbitrarily close to 1.

Proof (informal): We begin by constructing the piecewise-linear function suggested in the figure. ε is presumed to be very small, and ε' smaller still. The segment AB has slope

$$\frac{2\varepsilon'}{1-2\varepsilon} \approx 2\varepsilon'(1+2\varepsilon)$$
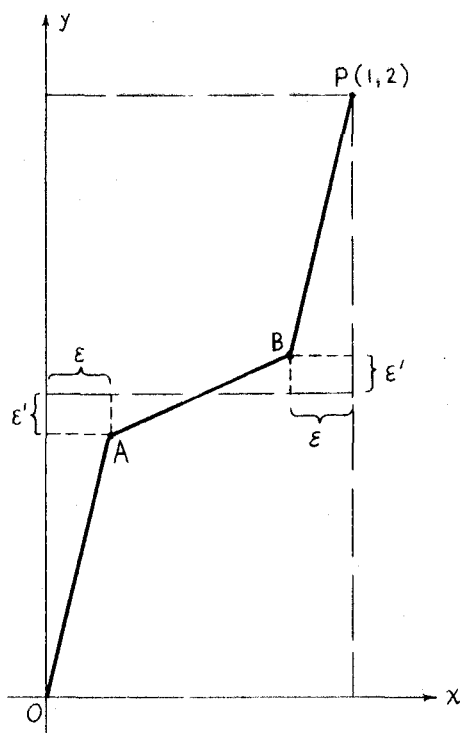
and OA and BP both have slope

$$\frac{1-\varepsilon'}{\varepsilon}.$$

The product of these slopes is

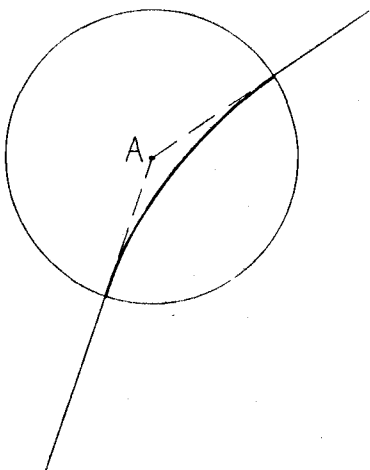$$\approx 2\frac{\varepsilon'}{\varepsilon}(1+2\varepsilon-\varepsilon')$$

and will be bigger than 1 if ε' is chosen so as to be at least half as big as ε; for

example, take $\varepsilon' = \frac{3}{4} \varepsilon$.

Now, in order to obtain an admissible function φ, describe a circle of radius r about points A and B; a smooth curve can be constructed to replace the segments falling within the circles so that the resulting function is $C^2[0,1]$ (in fact, $C^\infty[0,1]$ is attainable). The r here is at our disposal, and it shall be most convenient to think of it as being <u>infinitesimal</u> compared to $\varepsilon - \varepsilon'$; for then we may safely "reason" directly from the diagram on the previous page. We see that the subset V for this function is contained (substantially) within $[\varepsilon, 1-\varepsilon]$. If x lies in this range, then Tx will fall within one of the two extreme intervals where φ' is large. Thus we have now a function $\varphi \in \mathcal{A}_2$ which satisfies the requirements of the theorem, insofar as ε is in no way restricted as to degree of smallness.

PART 2

My goal in this section will be to pursue the relationship between the Rényi function, φ, and its invariant density, h. To that end the following theorem will prove vital.

<u>Theorem 8</u>: Let $\varphi \in \mathcal{R}$. Then the iterated distributions Φ_n converge uniformly, as $n \to \infty$, to the invariant distribution Φ.

In order to prove this result, I begin by showing that the transformation T is <u>mixing</u>, which means that for arbitrary Borel sets A, B in I,

$$\lim_{n \to \infty} \nu(T^{-n}A \cap B) = \nu(A)\nu(B).$$

(Recall that ν is the invariant measure; Lebesgue measure is denoted by μ.) Let $\mathcal{F}_n$ be the σ-algebra consisting of sets of the form $T^{-n}A$, A being a Borel set in I. The $\mathcal{F}_n$ form a decreasing sequence of σ-algebras; the limit

$$\mathcal{F} = \bigcap_{n=1}^{\infty} \mathcal{F}_n$$

is called the tail σ-algebra. It can be shown (see Billingsley [11], p. 121) that T is mixing if $\mathcal{F}$ contains

only sets of measure 0 or 1. I verify now that this is the case.

Let [a,b] be a subinterval of I, and let D_n denote the interval of rank n with left endpoint $[a_1, a_2, \ldots, a_n] = x$. Using the customary notation for conditional probability, we have

$$\mu(T^{-n}[a,b] \mid D_n) = \frac{[a_1, \cdots, a_n + b] - [a_1, \cdots, a_n + a]}{[a_1, \cdots, a_n + 1] - [a_1, \cdots, a_n]}$$

$$= \frac{H_n(x, t_1)}{H_n(x, t_2)} (b-a)$$

where $t_1 \in (a,b)$ and $t_2 \in (0,1)$. By Rényi's condition,

(35) $$\frac{b-a}{C} \leq \mu(T^{-n}[a,b] \mid D_n) \leq C(b-a).$$

It is clear that we can claim (35) with [a,b] replaced by an arbitrary Borel set A:

(36) $$\frac{\mu(A)}{C} \leq \mu(T^{-n}A \mid D_n) \leq C\mu(A).$$

Using (12), we may convert (36) to an inequality about ν, which works out to be

(37) $$\frac{\nu(A)}{C^4} \leq \nu(T^{-n}A \mid D_n) \leq C^4 \nu(A).$$

Suppose now that A is a set in the tail σ-algebra $\mathcal{Y}$. Then for any n, there is a Borel set B such that $A = T^{-n}B$. Thus from (37) we obtain

(38) $$\frac{\nu(A)}{C^4} = \frac{\nu(T^{-n}B)}{C^4} = \frac{\nu(B)}{C^4} \leq \nu(T^{-n}B \mid D_n) = \nu(A \mid D_n).$$

If $\nu(A) > 0$, then we may recast (38) in the form

(39) $\qquad \nu(D_n) = \dfrac{\nu(A)\,\nu(D_n|A)}{\nu(A|D_n)} \leq C^4 \nu(D_n|A).$

The fundamental intervals may be used to generate the Borel sets; therefore we may deduce from (39)

$$\nu(E) \leq C^4 \nu(E|A)$$

for an arbitrary Borel set E. Taking E to be the complement of A, we find that $\nu(E) = 0$, which is to say $\nu(A) = 1$. This shows that $\mathcal{G}$ contains only sets of measure 0 and 1, and hence that the transformation T is mixing.

We are now in a position to complete the proof of Theorem 8. I shall demonstrate that, for arbitrary measurable $A \subset I$,

(40) $\qquad \displaystyle\lim_{n \to \infty} \mu(T^{-n}A) = \nu(A)$

and to do this I shall use the equality

$$\mu(T^{-n}A) = \int_{T^{-n}A} d\mu = \int_{T^{-n}A} \frac{1}{h}\, d\nu.$$

Now, because T is mixing, we have

$$\nu(T^{-n}A \cap B) = \int_{T^{-n}A} I_B \, d\nu \longrightarrow \nu(A)\,\nu(B) \quad as \ n \to \infty,$$

I_B being the characteristic function of the set B. By taking linear combinations of such characteristic functions, we conclude that

(41) $$\lim_{n\to\infty} \int_{T^{-n}A} s\, d\nu = \nu(A) \int_I s\, d\nu$$

for an arbitrary simple function, s. Since 1/h is
measurable and bounded, we may approximate it <u>uniformly</u> by
simple functions. Therefore we may replace s in (41) by
1/h and write

$$\lim_{n\to\infty} \mu(T^{-n}A) = \lim_{n\to\infty} \int_{T^{-n}A} \frac{1}{h}\, d\nu = \nu(A) \int_I \frac{1}{h}\, d\nu = \nu(A),$$

proving (40). We take, for A, the interval [0,t] and
recall

$$\mu(T^{-n}[0,t]) = \Phi_n(t)$$

to conclude, finally,

$$\lim_{n\to\infty} \Phi_n(t) = \nu([0,t]) = \Phi(t).$$

That the convergence is uniform follows automatically from
the fact that all the Φ_n, and Φ as well, are both
continuous and monotone. This completes the proof.

The next three theorems will deal with the question of
how the behaviour of h is affected by various smoothness
conditions imposed on φ. The first two of these theorems
are dependent upon a curious additional restriction on φ,
which I refer to as <u>condition star</u>:

$$(*) \qquad \sup_{t\in I} \sum_{k=0}^{p-1} \{f'(k+t)\}^2 < 1.$$

The most obvious effect of this restriction is to force membership in $\mathcal{A}_1$, since $\varphi'(x)$ must clearly remain > 1. Aside from this, the condition seems mild; heuristically, we expect

$$\sum_{k=0}^{p-1} f'(k+t) \approx \int_0^p f'(t)\,dt = 1.$$

If none of the individual terms on the left hand side is especially close to 1, then the effect of squaring these terms should be to reduce the sum safely below 1. In particular, note that (*) will automatically obtain if $\sup_{x \in [0,p]} f'(x) < \frac{1}{\sqrt{p}}$; that is,

$$(42) \qquad \inf_{t \in I} \varphi'(t) > \sqrt{p} \implies (*).$$

__Theorem 9:__ Let $\varphi \in \mathcal{A}_1$ satisfy (*). Then h is continuous on I, and is the uniform limit of the iterated densities.

__Proof:__ Differentiating the recurrence formula (10) for the iterated densities gives

$$(43) \quad S_{n+1}'(t) = \sum_{k=0}^{p-1} \left\{ S_n'(f(k+t))\,(f'(k+t))^2 + S_n(f(k+t))\,f''(k+t) \right\}.$$

We know from the iterated densities lemma that the S_n are uniformly bounded by C. Put

$$B_n = \sup_{t \in I} \left| S_n'(t) \right|,$$

$$d = \sup_{t \in I} \sum_{k=0}^{p-1} \left| f''(k+t) \right|$$

and let $\theta < 1$ be the constant defined by the left hand side of (*). It follows now from (43) that

$$B_{n+1} \leq B_n \theta + C d.$$

From this we see at once that the B_n are themselves uniformly bounded by some fixed number B. Hence the iterated densities are uniformly bounded and equicontinuous. By the Ascoli theorem, there exists a uniformly convergent subsequence $\{\mathcal{S}_n(t)\}$ with continuous limit:

$$\lim_{n \to \infty} \mathcal{S}_n(t) = g(t), \quad \text{say.}$$

Since the convergence is uniform on I, we may integrate to obtain

$$\lim_{n \to \infty} \int_0^t \mathcal{S}_n(\tau) \, d\tau = G(t), \quad t \in I.$$

But $\left\{ \int_0^t \mathcal{S}_n(\tau) \, d\tau \right\}$ is a subsequence of $\{ \Phi_n(t) \}$, and so from Theorem 8 we conclude that $G(t) = \Phi(t)$, whence $G'(t) = g(t)$ must be equal to the invariant density $h(t)$. Thus h is continuous on I.

To show that the full sequence of iterated densities converges uniformly to h, assume the contrary. Then it must be possible to find an $\varepsilon > 0$ and a subsequence $\{\mathcal{S}_n\}$ of $\{S_n\}$ such that

$$(44) \qquad \sup_{t \in I} \left| \mathcal{S}_n(t) - h(t) \right| \geq \varepsilon \quad \text{for all } n.$$

The subsequence $\{\delta_n\}$ is itself uniformly bounded and equicontinuous, and so must possess a uniformly convergent subsequence $\{\sigma_n\}$ with limit, say g. As before, we find that g = h, and this is incompatible with (44). Thus the theorem is proved.

By an extension of the same reasoning one can see that h will be "nearly" as differentiable as φ: $\varphi \in C^n[0,1] \Rightarrow h \in C^{n-2}[0,1]$. In particular, we have

Theorem 10: Let $\varphi \in \mathcal{A}_1 \cap C^\infty[0,1]$ and satisfy (*). Then $h \in C^\infty[0,1]$. Furthermore, for $k \geq 0$, $S_n^{(k)}$ converges uniformly to $h^{(k)}$ as $n \to \infty$.

Proof: To begin with, differentiate (43):

$$(45) \qquad S_{n+1}'' = \sum_{k=0}^{p-1} \left\{ S_n'' \cdot (f')^3 + 3 S_n' \cdot (f'f'') + S_n \cdot f''' \right\}$$

and put

$$\alpha = \sup_{t \in I} \sum_{k=0}^{p-1} \left(f'(k+t) \right)^3 < \theta,$$

$$\beta = \sup_{t \in I} \sum_{k=0}^{p-1} 3 \left| f'(k+t) f''(k+t) \right|,$$

$$\text{and} \quad \gamma = \sup_{t \in I} \sum_{k=0}^{p-1} \left| f'''(k+t) \right|.$$

We already know the S_n to be uniformly bounded by C, and the $|S_n'|$ to be uniformly bounded by B. If we put

$$A_n = \sup_{t \in I} \left| S_n''(t) \right|$$

then we read from (45) that

$$A_{n+1} \leq A_n \alpha + B \beta + C \gamma.$$

The last two terms being constants independent of n, and $\alpha < \theta$ being smaller than 1, we perceive as before a uniform bound A for the A_n. Hence the sequence of first derivatives $\{S_n'\}$ is uniformly bounded and equicontinuous, and thus possesses a uniformly convergent subsequence $\{\mathcal{S}_n'\}$:

$$\lim_{n \to \infty} \mathcal{S}_n'(t) = g'(t), \quad \text{say.}$$

Integrating twice and using the same argument as before, we find that g = h so that h' = g' exists and is continuous. Again as before, we can show that the complete sequence $\{S_n'\}$ converges uniformly to h' as $n \to \infty$.

It is now apparent how the induction is to proceed, but having introduced constants in the order C, B, A, I am obliged to stop at this stage.

By way of illustration, let me now verify condition (*) for the function

$$(46) \qquad \varphi(x) = (1+x)^m - 1, \quad m \text{ an integer} \geq 2,$$

which is clearly $\in \mathcal{A}_1$. The inverse function is given by

$$f(x) = (1+x)^{\frac{1}{m}} - 1.$$

Thus,

$$\theta_m = \sup_{t \in I} \sum_{k=0}^{2^m-2} \left\{ f'(k+t) \right\}^2 = \sum_{k=0}^{2^m-2} \frac{1}{m^2} (k+1)^{\frac{2}{m}-2}$$

$$= \frac{1}{m^2} \sum_{k=1}^{2^m-1} \frac{1}{k^{2-\frac{2}{m}}} .$$

If m = 2, the series would diverge if not truncated. For m > 2, we can use the estimate

$$\theta_m < \frac{1}{m^2} \sum_{k=1}^{\infty} \frac{1}{k^{2-\frac{2}{m}}} .$$

The expression on the right decreases with increasing m, and therefore assumes its largest value when m = 3. Thus

$$\theta_m < \frac{1}{9} \sum_{k=1}^{\infty} \frac{1}{k^{4/3}} < \frac{1}{9} \left(1 + \int_1^{\infty} x^{-4/3} dx \right) = \frac{4}{9} < 1,$$

verifying (*) for m > 2. If m = 2, we have

$$\inf_{t \in I} \varphi'(t) = m = 2 ; \qquad \sqrt{p} = \sqrt{2^m - 1} = \sqrt{3} < 2,$$

thus fulfilling (*) according to (42). Therefore the invariant density h belonging to (46) is infinitely differentiable. The final theorem will show that it is even analytic.

Theorem 11: Let $\varphi \in \mathcal{A}_1$ be analytic on I. Then h is analytic on I, and moreover, for $k \geq 0$, $h^{(k)}$ is the uniform limit of $S_n^{(k)}$ as $n \to \infty$.

<u>Proof</u>: The idea is to show the existence of a region $\mathscr{D}$ in the complex plane, containing the unit interval I, such that the iterated distributions are analytic and uniformly bounded in $\mathscr{D}$. Since we are guaranteed convergence at least on I, all of our conclusions will follow from the Vitali convergence theorem (see Titchmarsh [13], p. 168).

For a given $\varepsilon > 0$ let $\mathscr{D}$ be the region containing I all of whose boundary points are at distance ε from I. Thus,

(47) $\quad t \in \mathscr{D} \iff t = t_r + \varepsilon \zeta \quad$ where $t_r \in I$ and $|\zeta| \leq 1.$

It is clearly possible to choose ε so that

$$\inf_{t \in \mathscr{D}} | \varphi'(t) | > 1,$$

and we do so.

Let $\mathscr{F}$ be the image of $\mathscr{D}$ under φ. Using the notation scheme of Theorem 1, we have

$$0 < \alpha = \inf_{t \in \mathscr{F}} |f'(t)| \leq \sup_{t \in \mathscr{F}} |f'(t)| = \beta < 1.$$

To show that all the Φ_n are defined and analytic on $\mathscr{D}$, we recall (9), according to which it is sufficient to show that

$$t \in \mathscr{D} \implies f(k+t) \in \mathscr{D} \quad for \ k = 0, 1, \cdots, p-1.$$

Using (47), we have

$$f(k+t) = f(k+t_r + \varepsilon \zeta) = f(k+t_r) + \varepsilon \zeta \, f'(k+t_r + \theta \varepsilon \zeta)$$

for some $\theta \in (0,1)$, and since $|\zeta f'| < 1$ we see that this is again a point in $\mathcal{D}$. This also establishes that expressions of the form

$$[\, a_1, a_2, \cdots, a_n + t \,]$$

are analytic for $t \in \mathcal{D}$. Define

$$\hat{\Delta}_n = \sup_{\substack{t_1, t_2 \in \mathcal{D} \\ a_1, a_2, \ldots, a_n}} \left| \, [\, a_1, a_2, \cdots, a_n + t_1 \,] - [\, a_1, a_2, \cdots, a_n + t_2 \,] \, \right|.$$

Thus,

$$\hat{\Delta}_1 = \sup_{\substack{t_1, t_2 \in \mathcal{D} \\ k}} \left| f(k+t_1) - f(k+t_2) \right| \leq K\beta$$

where $K = \operatorname{diam}(\mathcal{D}) = 1 + 2\varepsilon$. By a simple induction we find that

$$(48) \qquad \hat{\Delta}_n \leq K\beta^n \qquad \text{for } n \geq 1.$$

Using this we establish "Rényi's condition" throughout $\mathcal{D}$:

$$(49) \qquad \frac{\sup\limits_{t \in \mathcal{D}} \left| H_n(x,t) \right|}{\inf\limits_{t \in \mathcal{D}} \left| H_n(x,t) \right|} \leq C.$$

Thus, the sup and inf are assumed at two points t_1 and t_2 which happen to lie on the boundary of $\mathcal{D}$. We write everything out and, except for the modulus signs and the factor K in (48), all goes precisely as in the proof of

Theorem 1.

Next we extend the iterated densities lemma to show a uniform bound for $|S_n(t)|$, $t \in \mathscr{O}$. Using (11) and (49) we find

$$\sup_{t \in \mathscr{O}} |S_n(t)| = \sup_{t \in \mathscr{O}} \left| \sum_x H_n(x,t) \right| \leq \sup_{t \in \mathscr{O}} \sum_x |H_n(x,t)|$$

$$\leq \sum_x \sup_{t \in \mathscr{O}} |H_n(x,t)| \leq C \sum_x \inf_{t \in \mathscr{O}} |H_n(x,t)|$$

$$\leq C \sum_x \inf_{t \in I} |H_n(x,t)| = C \sum_x \inf_{t \in I} H_n(x,t)$$

$$\leq C \inf_{t \in I} \sum_x H_n(x,t) = C \inf_{t \in I} S_n(t) \leq C,$$

where, as before, the last step follows from $\int_0^1 S_n(t)dt = 1$.

Of course, it follows now that

$$|\Phi_n(t)| = \left| \int_0^t S_n(\tau)\,d\tau \right| \leq (1+\varepsilon) C$$

for all $t \in \mathscr{O}$, and the theorem is proved.

BIBLIOGRAPHY

1. Sôichi Kakeya, "On a generalized scale of notations", Japanese Journal of Mathematics, Vol 1 (1924) pp 95-108.

2. B.H. Bissinger, "A generalization of continued fractions", Bull. Amer. Math. Soc., 50 (1944), pp 868-876.

3. C.J. Everett, "Representations for real numbers", Bull. Amer. Math. Soc., 52 (1946), pp 861-869.

4. J.V. Uspensky, Introduction to Mathematical Probability, McGraw-Hill, New York (1937).

5. R.O. Kuz'min, "Sur un problème de Gauss", Atti de Congresso Internazionale de Mathematici Bologna, (1928), Vol VI, pp 83-89.

6. A. Ya. Khinchin, Continued Fractions, University of Chicago Press, Chicago (1964).

7. Paul Lévy, Théorie de l'Addition des Variables Aleatoires, Gauthier-Villars, Paris (1937).

8. W. Doeblin, "Remarques sur la théorie métrique des fractions continues", Compositio Math. 7, pp 353-371 (1940).

9. C. Ryll-Nardzewski, "On the ergodic theorems II", Studia Math. 12, pp 74-79 (1951).

10. Mark Kac, Statistical Independence in Probability,

Analysis and Number Theory, Carus Mathematical
Monographs 12, John Wiley & Sons, New York (1959).

11. Patrick Billingsley, Ergodic Theory and Information, John Wiley & Sons, New York (1965).

12. Alfréd Rényi, "Representations of real numbers and their ergodic properties", Acta. Math. Acad. Sci. Hungar., 8, pp 477-493 (1957).

13. E.C. Titchmarsh, The Theory of Functions, Oxford University Press, London (1939).

APPENDIX

APPENDIX: Proof of the assertion made on page 27.

Let $\varphi(x)$ be an admissible function whose derivative assumes the value 1 at the single point x = 0; moreover, suppose $\varphi''(x) > 0$ for x > 0. I wish to show that

$$\sum_{n=1}^{\infty} \Delta_n < \infty$$

provided

(50)
$$\int_{\varepsilon}^{1} \frac{x}{\varphi(x) - x} \, dx$$

remains finite as $\varepsilon \to 0$ from above.

To begin with, I show that the largest interval of rank n is the one lying farthest to the left; this is to say

(51)
$$\Delta_n = [\, 0, 0, \cdots, 0, 1 \,].$$

For convenience, I make a preliminary observation:

Lemma: Let us temporarily agree to say that a twice differentiable function is a sweetheart if its first derivative is positive and its second derivative is negative. Then if α and β are two sweethearts, so is the composite function $\alpha(\beta)$.

Proof: $\quad (\alpha(\beta))' = \alpha' \cdot \beta' > 0;$

$\quad\quad\quad (\alpha(\beta))'' = \alpha'' \cdot (\beta')^2 + \alpha' \cdot \beta'' < 0.$

Now, to establish (51), consider trying to maximize

(52) $\quad [a_1, a_2, \cdots, a_n+1] - [a_1, a_2, \cdots, a_n].$

Replace a_1 by the continuous variable t, and put

$$d_1 = [a_2, \cdots, a_n] < [a_2, \cdots, a_n+1] = d_2.$$

Then (52) can be written as

(53) $\quad f(t+d_2) - f(t+d_1).$

The derivative is

(54) $\quad f'(t+d_2) - f'(t+d_1) = f''(\theta)(d_2 - d_1)$

for some $\theta \in (t+d_1, t+d_2)$. Since $\varphi''(x) > 0$ for $x > 0$, we must have $f''(x) < 0$ for $x > 0$, so that the right hand side of (54) is strictly negative for $t \geq 0$. Thus (53) is largest when $t = 0$. This means we must take $a_1 = 0$ in (52), which leaves us trying to maximize

$$[0, a_2, \cdots, a_n+1] - [0, a_2, \cdots, a_n]$$

(55)
$$= f([a_2, \cdots, a_n+1]) - f([a_2, \cdots, a_n]).$$

As before, replace a_2 by the variable t, and put

$$d_1 = [a_3, \cdots, a_n] < [a_3, \cdots, a_n+1] = d_2.$$

Then (55) becomes

$$f(f(t+d_2)) - f(f(t+d_1)).$$

Here again we shall find a largest value for t = 0, for the simple reason that the composite function f(f) has, like f, a negative second derivative. Thus we will have to choose $a_2 = 0$. In view of our lemma, we may proceed inductively, arriving -- many sweethearts later -- at $a_1 = a_2 = \ldots = a_n = 0$. This establishes (51).

Let me now write f(x) in the form

(56) $$f(x) = x - \xi(x)$$

where $\xi(x)$ satisfies

(i) $\xi(0) = 0$

(ii) $\xi'(0) = 0$

(iii) $\xi'(x) > 0$ for $x > 0$

(iv) $\xi''(x) > 0$ for $x > 0$.
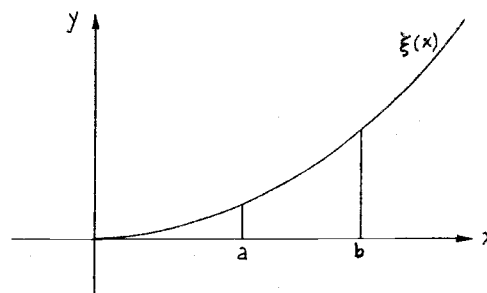
We note that $\triangle_{n+1} = f(\triangle_n) = \triangle_n - \xi(\triangle_n)$. Thus the number of terms among the $\triangle_n$ which lie in an interval [a,b] cannot exceed

$$\frac{b - a}{\xi(a)}$$

and the <u>sum</u> of these terms cannot exceed

$$b \cdot \frac{b-a}{\xi(a)} \, .$$

If we now break the interval $[a,b]$ into many subintervals $[x_{i-1}, x_i]$, and apply the same reasoning to each one of them, we find that the sum of those $\triangle_n$ lying in $[a,b]$ cannot exceed

$$\sum_i \frac{x_i}{\xi(x_{i-1})} \cdot (x_i - x_{i-1}) \, .$$

From this we easily extract the following criterion:
$\sum_{n=1}^{\infty} \triangle_n$ converges provided the integral

$$\int_a^b \frac{x}{\xi(x)} \, dx$$

remains finite as the lower limit approaches 0 from above. The equivalence of this criterion and the one based on the integral in (50) will follow from the like behaviour of the integrands near $x = 0$.

Writing $x = \varphi(t)$ in (56) gives

(57) $t = \varphi(t) - \xi(\varphi(t))$.

From the mean value theorem we have

$$\xi(\varphi(t)) = \xi(t) + \xi'(\theta) \cdot (\varphi(t) - t)$$

for some $\theta \in (t, \varphi(t))$. Substituting this in (57) and rearranging gives, finally,

$$\frac{\varphi(t) - t}{\xi(t)} = \frac{1}{1 - \xi'(\theta)} , \qquad \theta \in (t, \varphi(t)).$$

Since $\theta \to 0$ as $t \to 0$, we have

$$\lim_{t \to 0^+} \frac{\varphi(t) - t}{\xi(t)} = 1 ,$$

and the proof is complete.