Lectures on
Measure Theory and Probability

by

H.R. Pitt

Tata institute of Fundamental Research, Bombay
1958
(Reissued 1964)

Lectures on
Measure Theory and Probability

by

H.R. Pitt

Notes by

Raghavan Narasimhan

No part of this book may be reproduced in any form by print, microfilm or any other means without written permission from the Tata institute of Fundamental Research, Colaba, Bombay 5

Tata institute of Fundamental Research, Bombay
1957
(Reissued 1963)

Contents

Chapter 1

Measure Theory

1. Sets and operations on sets

We consider a space $\mathfrak{X}$ of elements (or point) x and systems of this subsets $X, Y, \ldots$ The basic relation between sets and the operations on them are defined as follows: 1

(a) *Inclusion:* We write $X \subset Y$ (or $Y \supset X$) if every point of X is contained in Y. Plainly, if 0 is empty set, $0 \subset X \subset \mathfrak{X}$ for every subset X. Moreover, $X \subset X$ and $X \subset Y$, $Y \subset Z$ imply $X \subset Z$. $X = Y$ if $X \subset Y$ and $Y \subset X$.

(b) *Complements:* The complements X' of X is the set of point of $\mathfrak{X}$ which do not belong to X. Then plainly $(X')' = X$ and $X' = Y$ if $Y' = X$. In particular , $O' = \mathfrak{X}$, $\mathfrak{X}' = 0$. Moreover, if $X \subset Y$, then $Y' \subset X'$.

(c) *Union:* The union of any system of sets is the set of points x which belong to at least one of them. The system need not be finite or even countable. The union of two sets X and Y is written $X \cup Y$, and obviously $X \cup Y = Y \cup X$. The union of a finite or countable sequence of sets $X_1, X_2, \ldots$ can be written $\displaystyle\bigcup_{n=1}^{\infty} X_n$.

(d) *Intersection:* The intersection of a system of sets of points which belong to every set of the system. For two sets it is written $X \cap Y$

(or $X.Y$) and for a sequence $\{X_n\}$, $\bigcap_{n=1}^{\infty} X_n$. Two sets are disjoint if their intersection is 0, a system of sets is disjoint if every pair of
2 sets of the system is. For disjoint system we write $X+Y$ for $X \cup Y$ and $\sum X_n$ for $\cup X_n$, this notation implying that the sets are disjoint.

(e) *Difference:* The difference $X.Y'$ or $X - Y$ between two X and Y is the sets of point of X which do not belong to Y. We shall use the notation $X - Y$ for the difference only if $Y \subset X$.

It is clear that the operations of taking unions and intersection are both commutative and associative. Also they are related t to the operation of taking complements by

$$X.X' = 0, X + X' = \mathfrak{X}, (X \cup Y)' = X', Y', (X.Y)' = X' \cup Y'.$$

More generally

$$(\cup X)' = \cap X', (\cap X)' = \cup X'.$$

The four operations defined above can be reduced to two in several different ways. For examples they can all be expressed in terms of unions and complements. In fact there is complete duality in the sense that any true proposition about sets remains true if we interchange

0	and	$\mathfrak{X}$
$\cup$	and	$\cap$
$\cap$	and	$\cup$
$\subset$	and	$\supset$

and leave = and ′ unchanged all through.

A countable union can be written as a sum by the formula

$$\bigcup_{n=1}^{\infty} X_n = X_1 + X_1'.X_2 + X_1'.X_2'.X_3 + \cdots$$

2. Sequence of sets

3 A sequence of sets $X_1, X_2, \ldots$ is *increasing* if

$$X_1 \subset X_2 \subset X_3 \subset \ldots$$

decreasing If

$$X_1 \supset X_2 \supset X_3 \supset \ldots$$

The *upper limit*, lim sup X_n of a sequence $\{X_n\}$ of sets is the set of points which belong to X_n for infinitely many n. The *lower limit*, $\liminf X_n$ is the set of points which belong to X_n for all but a finite number of n. It follows that $\liminf X_n \subset \limsup X_n$ and if $\limsup X_n = \liminf X_n = X$, X is called the *limit* of the sequence, which then *coverage* to X.

It is easy to show that

$$\liminf X_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} X_m$$

and that

$$\limsup X_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} X_m.$$

Then if $X_n \downarrow$,

$$\bigcap_{m=n}^{\infty} X_m = \bigcap_{m=1}^{\infty} X_m.\ \liminf X_n = \bigcap_{m=1}^{\infty} X_m,$$

$$\bigcup_{m=n}^{\infty} X_m = X_n, \limsup X_n = \bigcap_{n=1}^{\infty} X_n,$$

$$\lim X_n = \bigcap_{n=1}^{\infty} X_n,$$

and similarly if $X_n \uparrow$,

$$\lim X_n = \bigcup_{n=1}^{\infty} X_n.$$

3. Additive system of sets

4 A system of sets which contains $\mathfrak{X}$ and is closed under a finite number of complement and union operations is called a (*finitely*) *additive system or a field*. It follows from the duality principle that it is then closed under a finite number of intersection operations.

If an additive system is closed under a countable number of union and complement operations (and therefore under countable under inter sections), it is called a *completely additive system*, a *Borel system* or a *σ-field*.

It follows that any intersection (not necessarily countable) of additive or Borel system is a system of the same type. Moreover, the intersection of *all* additive (of Borel) systems containing a family of sets is a uniquely defined minimal additive (or Borel) system containing the given family. The existence of *at least one* Borel system containing a given family is trivial, since the system of *all* subsets of $\mathfrak{X}$ is a Borel system.

A construction of the actual minimal Borel system containing a given family of sets has been given by Hausdorff (Mengenlehre,1927, p.85).

Theorem 1. *Any given family of subsets of a space $\mathfrak{X}$ is contained in a unique minimal additive system S_0 and in a unique minimal Borel system S.*

Example of a finitely additive system: The family of rectangles $a_i \leq$
5 $x_i < b_i (i = 1, 2, ..., n)$ in R_n is not additive, but has a minimal additive
S_0 consisting of all "element ary figures" and their complements. An elementary figure is the union of a finite number of such rectangles.

The intersections of sets of an additive (or Borel) system with a fixed set(of the system) from an additive (or Borel) subsystem of the original one.

4. Set Functions

Functions con be defined on a system of sets to take values in any given space. If the space is an abelian group with the group operation called addition, one can define the additivity of the set function.

Thus, if μ is defined on an additive system of sets, μ is *additive* if

$$\mu\left(\sum X_n\right) = \sum \mu(X_n)$$

for any *finite* system of (disjoint) sets X_n.

In general we shall be concerned only with functions which take real values. We use the convention that the value $-\infty$ is excluded but that μ may take the value $+\infty$. It is obvious that $\mu(0) = 0$ if $\mu(X)$ is additive and finite for at least one X.

For a simple example of an additive set function we may take $\mu(X)$ to be the *volume* of X when X is an elementary figures in R_n.

If the additive property extends to countable system of sets, the function is called *completely* additive, and again we suppose that $\mu(X) \neq$ $-\infty$. Complete additive of μ can defined even if the field of X is only 6
finitely additive, provided that X_n and $\sum X_n$ belong to it.

Example of a completely additive function: $\mu(X)$ = number of elements (finite of infinite) in X for all subsets X of $\mathfrak{X}$

Examples of additive, but not completely additive functions:

1. $\mathfrak{X}$ is an infinite set,

$$\begin{aligned} \mu(X) &= 0 \text{ if } X \text{ is a finite subset of } \mathfrak{X} \\ &= \infty \text{ if } X \text{ is an infinite subset of } \mathfrak{X} \end{aligned}$$

   Let X be a countable set of elements $(x_1, x_2, \ldots)$ of $\mathfrak{X}$.

   Then

$$\mu(x_n) = 0, \sum \mu(x_n) = 0, \mu(X) = \infty.$$

2. $\mathfrak{X}$ is the interval $0 \leq x < 1$ and $\mu(X)$ is the sum of the lengths of finite sums of open or closed intervals *with closure in* $\mathfrak{X}$. These sets

together with $\mathfrak{X}$ from an additive system on which μ is additive but not completely additive if $\mu(\mathfrak{X}) = 2$.

A non-negative, completely additive function μ defined on a Borel system S of subsets of a set $\mathfrak{X}$ is called a *measure*. It is bounded (or finite) if $\mu(\mathfrak{X}) < \infty$. it is called a probability measure if $\mu(\mathfrak{X}) = 1$. The sets of the system S are called *measurable sets*.

5. Continuity of set functions

Definition . *A set function μ is said to be continuous, from below if $\mu(X_n) \to \mu(X)$ whenever $X_n \uparrow X$. It is continuous from above if $\mu(X_n) \to \mu(X)$ whenever $X_n \downarrow X$ and $\mu(X_{n_o}) < \infty$ for some n_0.*

7 It is continuous if it is continuous from above *and* below. Continuity at 0 means continuity from above at 0.

(For general ideas about limits of set functions when $\{X_n\}$ is not monotonic, see Hahn and Rosenthal, Set functions, Ch. I).

The relationship between additivity and complete additivity can be expressed in terms of continuity as follows.

Theorem 2. (a) *A completely additive function is continuous.*

(b) *Conversely, an additive function is completely additive if it is either continuous from below or finite and continuous at 0. (The system of sets on which μ is defined need only be finitely additive).*

Proof. (a) If $X_n \uparrow X$, we write

$$
\begin{aligned}
X &= X_1 + (X_2 - X_1) + (X_3 - X_2) + \cdots, \\
\mu(X) &= -\mu(X_1) + \mu(X_2 - X_1) + \cdots \\
&= \mu(X_1) + \lim_{N\to\infty} \sum_{n=2}^{N} \mu(X_n - X_{n-1}) \\
&= \lim_{N\to\infty} \mu(X_N).
\end{aligned}
$$

On the other hand, if $X_n \downarrow X$ and $\mu(X_{n_0}) < \infty$, we write

$$X_{n_0} = X + \sum_{n=n_0}^{\infty}(X_n - X_{n+1})$$

$$\mu(X_{n_0}) = \mu(X) + \sum_{n=n_0}^{\infty}\mu(X_n - X_{n+1}), \text{ and } \mu(X) = lim \mu(X_n)$$

as above since $\mu(X_{n_0}) < \infty$.

(b) First, if μ is additive and continuous from below, and **8**

$$Y = Y_1 + Y_2 + Y_3 + \cdots$$

we write

$$Y = \lim_{N\to\infty}\sum_{n=1}^{N} Y_n,$$

$$\mu(Y) = \lim_{N\to\infty}\mu\left(\sum_{n=1}^{N} Y_n\right), \text{ since } \sum_{n=1}^{N} Y_n \uparrow Y$$

$$= \lim_{N\to\infty}\sum_{n=1}^{N}\mu(Y_n)$$

by finite additivity, and therefore $\mu(Y) = \sum_{n=1}^{\infty}\mu(Y_n)$.

On the other hand, if μ is finite and continuous at 0, and $X = \sum_{n=1}^{\infty} X_n$, we write

$$\mu(X) = \mu\left(\sum_{n=1}^{N} X_n\right) + \mu\left(\sum_{n=N+1}^{\infty} X_n\right)$$

$$= \sum_{n=1}^{N}\mu(X_n) + \mu\left(\sum_{n=N+1}^{\infty} X_n\right), \text{ by finite additivity,}$$

since $\sum_{N+1}^{\infty} X_n \downarrow 0$ and has finite μ. □

Theorem 3 (Hahn-Jordan)**.** *Suppose that μ is completely additive in a Borel system S of subsets of a space $\mathfrak{X}$. Then we can write $\mathfrak{X} = \mathfrak{X}^+ + \mathfrak{X}^-$ (where $\mathfrak{X}^+$, $\mathfrak{X}^-$ belong to S and one may be empty) in such a way that*

1. $0 \leq \mu(X) \leq \mu(\mathfrak{X}^+) = M \leq \infty$ *for* $X \subset \mathfrak{X}^+$,

$-\infty < m = \mu(\mathfrak{X}^-) \leq \mu(X) \leq 0$ *for* $X \subset \mathfrak{X}^-$

while $m \leq \mu(X) \leq M$ *for all* X.

Corollary 1. *The upper and lower bounds M, m of $\mu(X)$ in S are attained for the sets $\mathfrak{X}^+$, $\mathfrak{X}^-$ respectively and $m > -\infty$.*

9 *Moreover, $M < \infty$ if $\mu(X)$ is finite for all X. In particular, a finite* measure *is bounded.*

Corollary 2. *If we write*

$$\mu^+(X) = \mu(X \cdot \mathfrak{X}^+), \mu^-(X) = \mu(X \cdot \mathfrak{X}^-)$$

we have

$$\mu(X) = \mu^+(X) + \mu^-(X), \mu^+(X) \geq 0, \mu^-(X) \leq 0$$
$$\mu^+(X) = \sup_{Y \subset X} \mu(Y), \mu^-(X) = \inf_{Y \subset X} \mu(Y).$$

If we write $\overline{\mu}(X) = \mu^+(X) - \overline{\mu}(X)$, we have also

$$|\mu(Y)| \leq \overline{\mu}(X) \text{ for all } Y \subset X.$$

It follows from the theorem and corollaries that an additive function can always be expressed as the difference of two measures, of which one is bounded (negative part here). From this point on, it is sufficient to consider only measures.

Proof of theorem 3. *[Hahn and Rosenthal, with modifications] We suppose that $m < 0$ for otherwise there is nothing to prove. Let A_n be defined so that $\mu(A_n) \to m$ and let $A = \bigcup_{n=1}^{\infty} A_n$. For every n, we write*

$$A = A_k + (A - A_k), A = \bigcap_{k=1}^{n} [A_k + (A - A_k)]$$

This can be expanded as the union of 2^n sets of the form $\bigcap_{k=1}^{n} A_k^*$, $A_k^* = A_k$ or $A - A_k$, and we write B_n for the sum of those for which $\mu < 0$. (If there is no such set, $B_n = 0$). Then, since A_n consists of disjoint sets which either belong to B_n or have $\mu \geq 0$, we get 10

$$\mu(A_n) \geq (B_n)$$

Since the part of B_{n+1} which does not belong to B_n consists of a finite number of disjoint sets of the form $\bigcap_{k=1}^{n+1} A_k^*$ for each of which $\mu < 0$,

$$\mu(B_n \cup B_{n+1}) = \mu(B_n) + \mu(B_{n+1}B_n') \leq \mu(B_n)$$

and similarly

$$\mu(B_n) \geq \mu(B_n \cup B_{n+1} \cup \ldots \cup B_{n'})$$

for any $n' > n$. By continuity from below, we can let $n' \to \infty$,

$$\mu(A_n) \geq \mu(B_n) \geq \mu\left(\bigcup_{k=n}^{\infty} B_k\right)$$

Let $\mathfrak{X}^- = \lim_{n\to\infty} \bigcup_{k=n}^{\infty} B_k$. Then

$$\mu(x^-) \leq \lim_{n\to\infty} \mu(A_n) = m,$$

and since $\mu(x^-) \geq m$ by definition of m, $\mu(x^-) = m$.

Now, if X is any subset of $\mathfrak{X}^-$ and $\mu(X) > 0$, we have

$$m = \mu(\mathfrak{X}^-) = \mu(X) + \mu(\mathfrak{X}^- - X) > \mu(\mathfrak{X}^- - X)$$

which contradicts the fact that m is $\inf_{Y \subset \mathfrak{X}} \mu(Y)$.

This proves (1) and the rest follows easily.

It is easy to prove that corollary 2 holds also for a completely additive function on a *finitely* additive system of sets, but $\sup \mu(X)$, $\inf \mu(X)$ are then not necessarily attained.

6. Extensions and contractions of additive functions

11 We get a contraction of an additive (or completely additive) function defined on a system by considering only its values on an function defined on a system by considering only its values on an additive subsystem. More important, we get an *extension* by embedding the system of sets in a larger system and defining a set function on the new system so that it takes the same values as before on the old system.

The basic problem in measure theory is to prove the existence of a measure with respect to which certain assigned sets are measurable and have assigned measures. The classical problem of defining a measure on the real line with respect to which every interval is measurable with measure equal to its length was solved by Borel and Lebesgue. We prove Kolmogoroff's theorem (due to Caratheodory in the case of R_n) about conditions under which an additive function on a finitely additive system S_0 can be extended to a measure in a Borel system containing S_0.

Theorem 4. *(a) If $\mu(I)$ is non-negative and additive on an additive system S_0 and if I_n are disjoint sets of S_0 with $I = \sum_{n=1}^{\infty} I_n$ also in S_0, then*

$$\sum_{n=1}^{\infty} \mu(I_n) \leq \mu(I).$$

(b) In order that $\mu(I)$ should be completely additive, it is sufficient that

$$\mu(I) \leq \sum_{n=1}^{\infty} \mu(I_n).$$

12 *(c) Moreover, if (I) is completely additive, this last inequality holds whether I_n are disjoint or not, provided that $I \subset \bigcup_{n=1}^{\infty} I_n$.*

Proof. (a) For any N,

$$\sum_{n=1}^{N} I_n, I - \sum_{n=1}^{N} I_n$$

belong to S_0 and do not overlap. Since their sum is I, we get

$$\mu(I) = \mu\left(\sum_{n=1}^{N} I_n\right) + \mu\left(I - \sum_{n=1}^{N} I_n\right)$$

$$\geq \mu\left(\sum_{n=1}^{N} I_n\right) = \sum_{n=1}^{N} \mu(I_n)$$

by finite additivity. Part (a) follows if we let $N \to \infty$ and (b) is a trivial consequence of the definition.

For (c), we write

$$\bigcup_{n=1}^{\infty} I_n = I_1 + I_2 \cdot I_1' + I_3 \cdot I_1' \cdot I_2' + \cdots$$

and then

$$\mu(I) \leq \mu[\cup_{n=1}^{\infty} I_n] = \mu(I_1) + \mu(I_2 \cdot I_1') + \cdots$$

$$\leq \mu(I_1) + \mu(I_2) + \cdots$$

□

7. Outer Measure

We define the out or measure of a set X with respect to a completely additive non-negative $\mu(I)$ defined on a additive system S_0 to be $\inf \sum \mu(I_n)$ for all sequences $\{I_n\}$ of sets of S_0 which cover X (that is, $X \subset \bigcup_{n=1}^{\infty}$). 13

Since any I of S_0 covers itself, its outer measure does not exceed $\mu(I)$. On the other hand it follows from Theorem 4(c) that

$$\mu(I) \leq \sum_{n=1}^{\infty} \mu(I_n)$$

for *every* sequence (I_n) covering I, and the inequality remains true if the right hand side is replaced by its lower bound, which is the outer measure of I. It follows that the outer measure of a set I of S_0 is $\mu(I)$, and there is therefore no contradiction if we use the same symbol $\mu(X)$ for the outer measure of every set X, whether in S_0 or not.

Theorem 5. *If* $X \subset \bigcup_{n=1}^{\infty} X_n$*, then*

$$\mu(X) \leq \sum_{n=1}^{\infty} \mu(X_n)$$

Proof. Let $\epsilon > 0$, $\sum_{n=1}^{\infty} \epsilon_n \leq \epsilon$. Then we can choose $I_{n\nu}$ from S_0 so that

$$X_n \subset \bigcup_{\nu=1}^{\infty} I_{n\nu}, \sum_{\nu=1}^{\infty} \mu(I_{n\nu}) \leq \mu(X_n) + \epsilon_n,$$

and then, since

$$X \subset \bigcup_{n=1}^{\infty} X_n \subset \bigcup_{n,\nu=1}^{\infty} I_{n\nu},$$

$$\begin{aligned} \mu(X) \leq \sum_{n=1}^{\infty} \sum_{\nu=1}^{\infty} \mu(I_{n\nu}) &\leq \sum_{n=1}^{\infty} (\mu(X_n) + \epsilon_n) \\ &\leq \sum_{n=1}^{\infty} \mu(X_n) + \epsilon, \end{aligned}$$

14 and we can let $\epsilon \to 0$. □

Definition of Measurable Sets.

We say that X is *measurable* with respect to the function μ if

$$\mu(PX) + \mu(P - PX) = \mu(P)$$

for every P with $\mu(P) < \infty$.

Theorem 6. *Every set I of S_o is measurable.*

Proof. If P is any set with $\mu(P) < \infty$, and $\epsilon > 0$, we can define I_n in S_0 so that

$$P \subset \bigcup_{n=1}^{\infty} I_n, \sum_{n=1}^{\infty} \mu(I_n) \leq \mu(P) + \epsilon$$

Then

$$PI \subset \bigcup_{n=1}^{\infty} I \cdot I_n, p - PI \subset \bigcup_{n=1}^{\infty} (I_n - II_n)$$

and since II_n and $I_n - II_n$ both belong to S_0,

$$\mu(PI) \leq \sum_{n=1}^{\infty} \mu(II_n), \mu(P - PI) \leq \sum_{n=1}^{\infty} \mu(I_n - II_n)$$

and

$$\begin{aligned} \mu(PI) + \mu(P - PI) &\leq \sum_{n=1}^{\infty} (\mu(II_n) + \mu(I_n - II_n)) \\ &= \sum_{n=1}^{\infty} \mu(I_n) \leq \mu(P) + \epsilon \end{aligned}$$

by additivity in S_0. Since ϵ is arbitrary, **15**

$$\mu(PI) + \mu(P - PI) \leq \mu(P)$$

as required.

We can now prove the fundamental theorem. □

Theorem 7 (Kolmogoroff-Caratheodory)**.** *If μ is a non-negative and completely additive set function in an additive system S_0, a measure can be defined in a Borel system S containing S_0 and taking the original value $\mu(I)$ for $I \in S_0$.*

Proof. It is sufficient to show that the measurable sets defined above form a Borel system and that the outer measure μ is completely additive on it.

If X is measurable, it follows from the definition of measurablility and the fact that

$$PX' = P - PX, P - PX' = PX,$$
$$\mu(PX') + \mu(P - PX) = \mu(PX) + \mu(P - PX)$$

that X' is also measurable.

Next suppose that X_1, X_2 are measurable. Then if $\mu(P) < \infty$,

$$\begin{aligned}\mu(P) &= \mu(PX_1) + \mu(P - PX_1) \text{ since } X_1 \text{ is measurable}\\ &= \mu(PX_1X_2) + \mu(PX_1 - PX_1X_2) + \mu(PX_2 - PX_1X_2)\\ &\qquad + \mu(P - P(X_1 \cup X_2)) \text{ since } X_2 \text{ is measurable}\end{aligned}$$

Then, since

$$(PX_1 - PX_1X_2) + (PX_2 - PX_1X_2) + (P - P(X_1 \cup X_2)) = P - PX_1X_2,$$

16 it follows from Theorem 5 that

$$\mu(P) \geq \mu(PX_1X_2) + \mu(P - PX_1X_2)$$

and so X_1X_2 is measurable.

It follows at once now that the sum and difference of two measurable sets are measurable and if we take $P = X_1 + X_2$ in the formula defining measurablility of X_1, it follows that

$$\mu(X_1 + X_2) = \mu(X_1) + \mu(X_2)$$

When X_1 and X_2 are measurable and $X_1X_2 = 0$. This shows that the measurable sets form an additive system S in which $\mu(X)$ is additive. After Theorems 4(b) and 5, $\mu(X)$ is also completely additive in S. To complete the proof, therefore, it is sufficient to prove that $X = \bigcup_{n=1}^{\infty} X_n$ is measurable if the X_n are measurable and it is sufficient to prove this in the case of disjoint X_n.

If $\mu(P) < \infty$,

$$\mu(P) = \mu\left(P \sum_{n=1}^{n} X_n\right) + \mu\left(P - P \sum_{n=1}^{N} X_n\right)$$

since $\sum_{n=1}^{N} X_n$ is measurable,

$$\geq \mu\left(P \sum_{n=1}^{N} X_n\right) + \mu(P - PX) = \sum_{n=1}^{N} \mu(PX_n) + \mu(P - PX)$$

by definition of measurablility applied $N-1$ times, the X_n being disjoint.

Since this holds for all N, **17**

$$\mu(P) \geq \sum_{n=1}^{\infty} \mu(PX_n) + \mu(P - PX)$$
$$\geq \mu(PX) + \mu(P - PX),$$

by Theorem 5, and therefore X is measurable. □

Definition. *A measure is said to be* complete *if every subset of a measurable set of zero measure is also measurable (and therefore has measure zero).*

Theorem 8. *The measure defined by Theorem 7 is complete.*

Proof. If X is a subset of a measurable set of measure 0, then $\mu(X) = 0$, $\mu(PX) = 0$, and

$$\mu(P) \leq \mu(PX) + \mu(P - PX) = \mu(P - PX) \leq \mu(P),$$
$$\mu(P) = \mu(P - PX) = \mu(P - PX) + \mu(PX),$$

and so X is measurable.

The measure defined in Theorem 7 is not generally the minimal measure generated by μ, and the minimal measure is generally not complete. However, any measure can be completed by adding to the system of measurable sets (X) the sets $X \cup N$ where N is a subset of a set of measure zero and defining $\mu(X \cup N) = \mu(X)$. This is consistent with the original definition and gives us a measure since countable unions of sets $X \cup N$ are sets of the same form, $(X \cup N)' = X' \cap N' = X' \cap (Y' \cup N \cdot Y')$ (where $N \subset Y$, Y being measurable and of 0 measure) $= X_1 \cup N_1$ is of the same form and μ is clearly completely additive on this extended system.

The essential property of a measure is complete additivity or the equivalent continuity conditions of Theorem 2(a). Thus, if $X_n \downarrow X$ or 18
$X_n \uparrow X$, then $\mu(X_n) \to \mu(X)$, if $X_n \downarrow 0$, $\mu(X_n) \to 0$ and if $X = \sum\limits_1^\infty X_n$, $\mu(X) = \sum\limits_1^\infty \mu(X_n)$. In particular, the union of a sequence of sets of measure zero also has measure zero. □

8. Classical Lebesgue and Stieltjes measures

The fundamental problem in measure theory is, as we have remarked already, to prove the existence of a measure taking assigned values on a given system of sets. The classical problem solved by Lebesgue is that of defining a measure on sets of points on a line in such a way that every interval is measurable and has measure equal to its length. We consider this, and generalizations of it, in the light of the preceding abstract theory.

It is no more complicated to consider measures in Euclidean space R_K than in R_1. A set of points defined by inequalities of the form

$$a_i \leq x_i < b_i (i = 1, 2, \ldots, k)$$

will be called a *rectangle* and the union of a finite number of rectangles, which we have called an *elementary figure*, will be called simply a *figure*. It is easy to see that the system of figures and complements of figures forms a finitely additive system in R_k. The *volume* of the rectangle defined above is defined to be $\prod\limits_{i=1}^{k}(b_i - a_i)$. A figure can be decomposed into disjoint rectangles in many different ways, but it is easy to verify
19 that the sum of the volumes of its components remains the same, however, the decomposition is carried out. It is sufficient to show that this is true when one rectangle is decomposed to be $+\infty$, it is easy to show by the same argument that the volume function $\mu(I)$ is finitely additive on the system S_0 of figures and their complements.

Theorem 9. *The function* $\mu(I)$ *(defined above) is* completely *additive in* S_0.

Proof. As in Theorem 2, it is sufficient to show that if $\{I_n\}$ is a decreasing sequence of figures and $I_n \to 0$, then $\mu(I_n) \to 0$. If $\mu(I_n)$ does not $\to 0$, we can define $\delta > 0$ so that $\mu(I_n) \geq \delta$ for all n and we can define a decreasing sequence of figures H_n such that *closure* $\overline{H}_n$ of H_n lies in I_n, while

$$\mu(I_n - H_n) < \frac{\delta}{2}$$

It follows that $\mu(H_n) = \mu(I_n) - \mu(I_n - H_n) > \frac{\delta}{2}$ so that H_n, and therefore $\overline{H}_n$, contains at least one point. But the intersection of a decreasing sequence of non-empty closed sets $(\overline{H}_n)$ is non-empty, and therefore the H_n and hence the I_n have a common point, which is impossible since $I_n \downarrow 0$. □

The measure now defined by Theorem 7 is Lebesgue Measure.

9. Borel sets and Borel measure

The sets of the *minimal* Borel system which contains all figures are called Borel sets and the measure which is defined by Theorem 9 and 7 is called Borel measure when it is restricted to these sets. The following results follow immediately.

Theorem 10. *A sequence of points in R_K is Borel measurable and has* 20
measure 0.

Theorem 11. *Open and closed sets in R_K are Borel sets.*

(An open set is the sum of a sequence of rectangles, and a closed set is the complement of an open set).

Theorem 12. *If X is any (Lebesgue) measurable set, and $\epsilon > 0$, we can find an open set G and a closed set F such that*

$$F \subset X \subset G, \mu(G - P) < \in$$

Moreover, we can find Borel sets A, B so that

$$A \subset X \subset B, \mu(B - A) = 0.$$

Conversely, any set X for which either of these is true is measurable.

Proof. First suppose that X is bounded, so that we can find a sequence of rectangles I_n so that

$$X \subset \bigcup_{n=1}^{\infty} I_n, \sum_{n=1}^{\infty} \mu(I_n) < \mu(X) + \epsilon/4.$$

Each rectangle I_n can be enclosed in an open rectangle (that is, a point set defined by inequalities of the from $a_i < x_i < b_i, i = 1, 2, \ldots, k$, its measure is defined to be $\prod_{i=1}^{k}(b_i - a_i)Q_n$ of measure not greater than $\mu(I_n) + \dfrac{\epsilon}{2^n + 2}$. □

Then

$$X \subset Q = \bigcup_{n=1}^{\infty} Q_n, \mu(Q) \leq \sum_{n=1}^{\infty} \mu(Q_n) \leq \sum_{n=1}^{\infty} \mu(I_n) + \epsilon \sum_{n=1}^{\infty} \frac{1}{2^n + 2} \leq \mu(X) + \frac{\epsilon}{2}$$

Then Q is open and $\mu(Q - X) \leq \epsilon/2$.

21 Now any set X is the sum of a sequence of *bounded* sets X_n (which are measurable if X is), and we can apply this each X_n with $6/2^{n+1}$ instead of $\in$. Then

$$X = \sum_{n=1}^{\infty} X_n, X_n \subset Q_n, \sum_{n=1}^{\infty} Q_n = G,$$

where G is open and

$$G - X \subset \bigcup_{n=1}^{\infty} (Q_n - X_n), \mu(G - X) \leq \sum_{n=1}^{\infty} \mu(Q_n - X_n) \leq \sum_{n=1}^{\infty} \frac{\in}{2^n + 1} = \frac{\in}{2}$$

The closed set F is found by repeating the argument on X and complementing.

Finally, if we set $\in_n \downarrow 0$ and G_n, F_n are open and closed respectively,

$$F_n \subset X \subset G_n,\ \mu(G_n - F_n) < \in_n$$

and we put

$$A = \bigcup_{n=1}^{\infty} F_n,\ B = \bigcup_{n=1}^{\infty} G_n,$$

we see that

$$A \subset X \subset B, \mu(B - A) \leq \mu(G_n - F_n) \leq \in_n \text{ for all n},$$

and so

$$\mu(B - A) = 0,$$

while A, B are obviously Borel sets.

Conversely, if $\mu(P) < \infty$ and

$$F \subset X \subset G,$$

We have, since a closed set is measurable, **22**

$$\begin{aligned} \mu(P) &= \mu(PF) + \mu(P - PF) \\ &\geq \mu(PX) - \mu(P(X - F)) + \mu(P - PX) \\ &\geq \mu(PX) + \mu(P - PX) - \mu(X - F) \\ &\geq \mu(PX) + \mu(P - PX) - \mu(G - F) \\ &\geq \mu(PX) + \mu(P - PX) - \in \end{aligned}$$

true for every $\epsilon > 0$ and therefore

$$\mu(P) \geq \mu(PX) + \mu(P - PX)$$

so that X is measurable.

In the second case, X is the sum of A and a subset of B contained in a Borel set of measure zero and is therefore Lebesgue measurable by the completeness of Lebesgue measure.

It is possible to defined measures on the Borel sets in R_k in which the measure of a rectangle is not equal to its volume. All that is necessary is that they should be completely additive on figures. Measures of this kind are usually called positive *Stiltjes measures* in R_k and Theorems 11 and 12 remain valid for them but *Theorem 10 does not*. For example, a single point may have positive Stieltjes measure.

A particularly important case is $k = 1$, when a Stieltjes measure can be defined on the real line by any monotonic increasing function $\Psi(X)$. The figures I are finite sums of intervals $a_i \leq x < b_i$ and $\mu(I)$ is defined by

$$\mu(I) = \sum_i \{\Psi(b_i - 0) - \Psi(a_i - 0)\}.$$

The proof of Theorem 9 in this case is still valid. We observe that since $\underset{\beta \to b-0}{lim} \Psi(\beta) = \Psi(b-0)$, it is possible to choose β so that $\beta < b$ and $\Psi(\beta - 0) - \Psi(a-0) > \frac{1}{2}, [\Psi(b-0) - \Psi(a-0)]$. 23

The set function μ can be defined in this way even if $\Psi(x)$ is not monotonic. If μ is bounded, we say that $\psi(x)$ is of *bounded variation*. In this case, the argument of Theorem 9 can still be used to prove that μ is completely additive on figures. After the remark on corollary 2 of Theorem 3, we see that it can be expressed as the difference of two completely additive, non-negative functions $\mu^+, -\mu^-$ defined on figures. These can be extended to a Borel system of sets X, and the set function $\mu = \mu^+ + \mu^-$ gives a set function associated with $\Psi(x)$. We can also write $\Psi(x) = \Psi^+(x) + \Psi^-(x)$ where Ψ^+(x) increases, Ψ (x) decreases and both are bounded if $\Psi(x)$ has bounded variation.

A non-decreasing function $\Psi(x)$ for which $\Psi(-\infty) = 0, \Psi(\infty) = 1$ is called a *distribution function*, and is of basic importance in probability.

10. Measurable functions

A function $f(x)$ defined in $\mathfrak{X}$ and taking real values is called *measurable with respect to a measure* μ if $\varepsilon[f(x) \geq k]$($\varepsilon[P(x)]$ is the set of points x in $\mathfrak{X}$ for which $P(x)$ is true) is measurable with respect to μ for every real k.

Theorem 13. *The memorability condition*

24 (i) $\varepsilon[f(x) \geq k]$ is measurable for all real k is equivalent to each one of

(ii) $\varepsilon[f(x) > k]$ is measurable for all real k,

(iii) $\varepsilon[f(x) \leq k]$ is measurable for all real k,

(iv) $\varepsilon[f(x) < k]$ is measurable for all real k,

Proof. Since

$$\varepsilon[f(x) \geq k] = \bigcap_{n=1}^{\infty} \varepsilon\left[f(x) > k - \frac{1}{n}\right],$$

(ii) implies (i). Also

$$\varepsilon[f(x) \geq k] = \bigcup_{n=1}^{\infty} \varepsilon\left[f(x) \geq k + \frac{1}{n}\right],$$

and so (i)implies (ii). This proves the theorem since (i) is equivalent with (iv) and (ii) with (iii) because the corresponding sets are complements. □

Theorem 14. *The function which is constant in $\mathfrak{X}$ is measurable. If f and g are measurable, so are $f \pm g$ and $f \cdot g$.*

Proof. The first is obvious. To prove the second , suppose f, g are measurable. Then

$$\begin{aligned}\varepsilon[f(x) + g(x) > k] &= \varepsilon[f(x) > k - g(x)] \\ &= \cup\varepsilon[f(x) > r > k - g(x)] \\ &= \bigcup_{r}^{r} \varepsilon[f(x) > r] \cap \varepsilon[g(x) > k - r]\end{aligned}$$

the union being over all rationals r. This is a countable union of measurable sets so that $f + g$ is measurable. Similarly $f - g$ is measurable. Finally

$$\varepsilon[f(x))^2 > k] = \varepsilon[f(x) > \sqrt{k}] + \varepsilon[f(x) < -\sqrt{k}] \text{ for } k \geq 0$$

so that f^2 is measurable. Since 25

$$f(x)g(x) = \frac{1}{4}(f(x) + g(x))^2 - \frac{1}{4}(f(x) - g(x))^2$$

$f \cdot g$ is measurable. □

Theorem 15. *If f_n measurable for $n = 1, 2, \ldots$ then so are* $\limsup f_n$, $\liminf f_n$.

Proof. $\epsilon[\limsup f_n(x) < k]$

$$= \epsilon[f_n(x) < k \text{ for all sufficiently large } n]$$

$$= \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \epsilon[f_n(x) < k]$$

is measurable for all real k. Similarly $\liminf f_n$ is measurable. □

In R_n, a function for which $\epsilon[f(x) \geq k]$ is Borel measurable for all k is called a *Borel measurable function* or a *Baire function*.

Theorem 16. *In R_n, a continuous function is Borel measurable.*

Proof. The set $\epsilon[f(x) \geq k]$ is closed. □

Theorem 17. *A Baire function of a measurable function is measurable.*

Proof. The Baire functions form the smallest class which contains continuous functions and is closed under limit operations. Since the class of measurable functions is closed under limit operations, it is sufficient to prove that a continuous function of a measurable, function is measurable. Then if $\varphi(u)$ is continuous and $f(x)$ measurable, $\epsilon[\varphi(f(x)) > k]$
26 is the set of x for which $f(x)$ lies in an open Set, namely the open set of points for which $\varphi(u) > k$. Since an open set is a countable union of open intervals, this set is measurable, thus proving the theorem. □

Theorem 18 (Egoroff)**.** *If $\mu(X) < \infty$ and $f_n(x) \to f(x) \neq \pm\infty$ p.p in X, and if $\delta > 0$, then we can find a subset $X_\circ$ of X such that $\mu(X - X_\circ) < \delta$ and $f_n(x) \to f(x)$ uniformly in $X_\circ$.*

We write p.p for "almost everywhere", that is, everywhere expect for a set of measure zero.

Proof. We may plainly neglect the set of zero measure in which $f_n(X)$ dose not converge to a finite limit. Let

$$X_{N,\nu} = \epsilon[| f(x) - f_n(x) |< 1/\nu \text{ for all } \geq N].$$

Then, for fixed ν,

$$X_{N,\nu} \uparrow X \text{ as } N \to \infty$$

For each ν we choose N_ν os that $X_\nu = X_{N_\nu,\nu}$ satisfies

$$\mu(X - X_\nu) < \delta/2^\nu,$$

and let
$$X_\circ = \bigcap_{\nu=1}^{\infty} X_\nu$$

Then
$$\mu(X - X_\circ) \leq \sum_{\nu=1}^{\infty} \mu(X - X_\nu) < \delta$$

and
$$| f(x) - f_n(x) |< 1/\nu \text{ for } n \geq N_\nu$$

if x is in x_ν and therefore if X is in $X_\circ$. This proves the theorem. □

11. The Lebesgue integral

Suppose that $f(x) \geq 0$, f is measurable in X, and let **27**

$$0 = y_\circ < y_1 < y_2 \cdots < y_\nu \to \infty$$

and
$$E_\nu = \epsilon\left[y_\nu \geq f(x) < y_{\nu+1}\right], \nu = 0, 1, 2, \ldots$$

so that E is measurable and $x = \sum_{\nu=0}^{\infty} E_\nu$.

We call the set of the y_ν, $\{y_\nu\}$ subdivision.

Let
$$S = S\{y\} = \sum_{\nu=1}^{\infty} y_\nu \mu(E_\nu).$$

Then we define $\sup S$ for all subdivisions $\{y_\nu\}$ to be the *Lebesgue Integral* of $f(x)$ over X, and write it $\int f(X)d\mu$. We say that $f(x)$ is *integrable* or *summable* if its integral is finite. It is obvious that changes in the values of f in a null set (set of measure 0) have no effect on the integral.

Theorem 19. *Let $\{y_\nu^k\}, k = 1, 2, \ldots$, be a sequence of subdivisions whose maximum intervals*

$$\delta_k = \sup(y_{\nu+1}^k - y_\nu^k) \to 0 \text{ as } k \to \infty$$

Then, if S_k is the sum corresponding to $\{y_\nu^k\}$,

$$\lim_{k\to\infty} S_k = \int_{\mathfrak{X}} f(x)d\mu = F(\mathfrak{X}).$$

Corollary. *Since S_k is the integral of the function taking constant values y^k_ν in the sets E^k_ν, it follows, by leaving out suitable remainders $\sum_{\nu=v+1}^{\infty} y^k_\nu \mu(E^k_\nu)$, that $F(\mathfrak{X})$ is the limit of the integrals of simple func-*
28 *tions, a simple function being a function taking constant values on each of a finite number of measurable sets whose union is $\mathfrak{X}$.*

Proof. If $A < F(\mathfrak{X})$, we can choose a subdivision $\{y'_\nu\}$ so that if E_ν are the corresponding sets, S' the corresponding sum,

$$S' \geq \sum_{\nu=1}^{V} y'_\nu \mu(E'_\nu)$$

for a finite V. One of the $\mu(E'_\nu)$ can be infinite only if $F(x) = \infty$ and then there is nothing to prove. Otherwise, $\mu(E'_\nu) < \infty$ and we let $\{y_\nu\}$ be a subdivision with $\delta = \sup(y_{\nu+1} - y_\nu)$ and denote by S'' the sum defined for $\{y_\nu\}$ and by S the sum defined for the subdivision consisting of points y_ν and y'_ν. Since S' is not decreased by insertion of extra points of sub-division,

$$S'' \geq S' \geq \sum_{\nu=1}^{V} y'_\nu \mu(E'_\nu) > A,$$

while

$$S'' - S \leq \delta \sum_{1}^{V} \mu(E'_\nu)$$

and, by making δ small enough we get $S > A$. Since $S \leq F(\mathfrak{X})$ and $A < F(\mathfrak{X})$ is arbitrary, this proves the theorem. □

The definition can be extended to integrals over subsets X of by defining

$$F(X) = \int_X f(x) d\mu = \int_{\mathfrak{X}} f_x(X) d_\mu$$

where $f_X(x) = f(x)$ for x in x and $f_X(x) = 0$ for x in $\mathfrak{X} - X$. We may therefore always assume (when it is convenient) that integrals are over the whole space $\mathfrak{X}$.

29 The condition $f(X) \geq 0$ can easily be removed.

We define

$$f^+(x) = f(x) \text{ when } f(x) \geq 0, f^+(x) = 0 \text{ when } f(x) \leq 0,$$
$$f^-(x) = f(x) \text{ when } f(x) \leq 0, f^-(x) = 0 \text{ when } f(x) \geq 0.$$

Then $f(x) = f^+(x) + f^-(x), | f(x) |= f^+(x) - (x)$.
We define

$$\int_X f(x)d\mu = \int_X f^+(x)d\mu - \int_X (-f^-(x))d\mu$$

when both the integrals on the right are finite, so that $f(x)$ is integrable if and only if $| f(x) |$ is integrable.

In general, we use the integral sign only when the integrand is integrable in this absolute sense. The only exception to this rule is that we may sometimes write $\int_X f(x)d\mu = \infty$ when $f(x) \geq -r(x)$ and $r(x)$ is integrable.

Theorem 20. *If $f(x)$ is integrable on $\mathfrak{X}$, then*

$$F(X) = \int_X f(x)d\mu$$

is defined for every measurable subset X of $\mathfrak{X}$ and is completely additive on these sets.

Corollary. *If $f(x) \geq 0$, then $F(Y) \leq F(X)$ if $Y \subset X$*

Proof. It is sufficient to prove the theorem in the case $f(X) \geq 0$. Let $X = \sum_{n=1}^{\infty} X_n$ where are X_n are measurable and disjoint. Then, if $\{y_\nu\}$ is a subdivision, $E_\nu = \sum_{n=1}^{\infty} E_\nu X_n, \mu(E_\nu) = \sum_{n=1}^{\infty} \mu(E_\nu X_n)$ and

$$S = \sum_{\nu=1}^{\infty} y_\nu \mu(E_\nu) = \sum_{\nu=1}^{\infty} y_\nu \sum_{n=1}^{\infty} \mu(E_\nu X_n)$$

$$= \sum_{n=1}^{\infty} \sum_{\nu=1}^{\infty} y_{\nu}\mu\,(E_{\nu}X_n)$$
$$= \sum_{n=1}^{\infty} S_n$$

30 where S_n is the sum for $f(x)$ over X_n. Since S and S_n (which are ≥ 0) tend to $F(X)$ and $F(X_N)$ respectively as the maximum interval of subdivision tends to 0, we get

$$F(X) = \int_X f(x)d\mu = \sum_{n=1}^{\infty} F(X_n).$$

□

Theorem 21. *If a is a constant,*

$$\int_X af(x)d\mu = a\int_X f(x)d\mu$$

Proof. We may again suppose that $f(x) \geq 0$ and that $a > 0$. If we use the subdivision $\{y_{\nu}\}$ for $f(x)$ and $\{ay_{\nu}\}$ for af (x), the sets E_{ν} are the same in each case, and the proof is trivial. □

Theorem 22. *If $A \leq f(x) \leq B$ in X, then*

$$A\mu(X) \leq F(X) \leq B\mu(X).$$

Theorem 23. *If $f(x) \geq g(x)$ in X, then*

$$\int_X f(x)d\mu \geq \int_X g(x)d\mu$$

Corollary. *If $|f(x)| \leq g(x)$ and $g(x)$ is integrable, then so is $f(x)$.*

Theorem 24. *If $f(x) \geq 0$ and $\int_X f(x)d\mu = 0$, then $f(x) = 0$ p.p. in X.*

31 *Proof.* If this were not so, then

$$\in [f(x) > 0] = \sum_{n=0}^{\infty} \in \left[\frac{1}{n+1} \leq f(x) < \frac{1}{n}\right]$$

has positive measure, and hence, so has at least one subset E_n $= \in \left[\frac{1}{n+1} \leq f(x) < \frac{1}{n}\right]$ Then

$$\int_x f(x)d \geq \int_{E_n} f(x)d\mu \geq \frac{\mu(E_n)}{n+1} > 0$$

which is impossible. □

Corollary 1. *If* $\int_K f(x)d\mu = 0$ *for all* $X \subset \mathfrak{X}$, $f(x)$ *not necessarily of the same sign, then* $f(x) = 0$ *p.p.*

we have merely to apply Theorem 24 to $X_1 = \in [f(x) \geq 0]$ *and to* $X_2 = \in [f(x) < 0]$.

Corollary 2. *If* $\int_X f(x)d\mu = \int_x g(x)d\mu$ *for all* $X \subset \mathfrak{X}$ *, then* $f(x) = g(x)$ *p.p. If* $f(x) = g(x)$ *p.p. we say that f and g are equivalent.*

12. Absolute Continuity

A completely additive set function $F(x)$ defined on a Borel system is said to be *absolutely continuous* with respect to a measure μ on the same system if $F(X) \to o$ uniformly in X as $\mu(X) \to 0$. In other words, if $\in > 0$, we can find $\delta > 0$ so that $|F(X)| < \in$ for all sets X which satisfy $\mu(X) < \delta$. In particular, if $F(X)$ is defined in R by a point function $F(x)$ of bounded variation, then it is absolutely continuous, if given $\epsilon > o$ we can find $\delta > 0$ so that

$$\sum_{i=1}^{n} \left|F(b_i) - F(a_i)\right| \leq \in \text{ if } \sum_{i-1}^{n} (b_i - a_1) < \delta$$

32

Moreover, it is clear from the proof of Theorem 3 that a set function $F(X)$ is absolutely continuous if and only if its components $F^+(X)$, $F^-(X)$ are both absolutely continuous. An absolutely continuous point function $F(x)$ can be expressed as the difference of two absolutely continuous non-decreasing functions as we see by applying the method used on page 22 to decompose a function of bounded variation into two monotonic functions. We observe that the concept of absolute continuity does not involve any topological assumptions on X.

Theorem 25. *If $f(x)$ is integrable on X, then*

$$F(X) = \int_x f(x) d\mu$$

is absolutely continuous.

Proof. We may suppose that $f(x) \geq 0$. If $\in > 0$, we choose a subdivision { y_ν }so that

$$\sum_{\nu=1}^{\infty} y_\nu \mu(E_\nu) > F(\mathfrak{X}) - \in /4$$

and then choose V so that

$$\sum_{\nu=1}^{V} y_\nu \mu(E_\nu) > F(\mathfrak{X}) - \in /2$$

Then, if $A > y_{\nu+1}$ and $E_A = \varepsilon[f(x) \geq A]$

we have $$E_\nu \subset \mathfrak{X} - E_A \text{ for } \nu \leq V.$$

Now $$F(\mathfrak{X} - E_A) \geq \sum_{\nu=1}^{V} \int_{E_V} f(x) d\mu \geq \sum_{\nu=1}^{V} y_\nu \mu(E_\nu)$$
$$> F(\mathfrak{X}) - \in /2$$

and therefore, $$F(E_A) < \in /2.$$

33 If X is any measurable set,

$$F(x) = F(XE_A) + F(X - E_A)$$
$$< \frac{\in}{2} + A\mu(X)(\text{since} f(x) \leq A \text{ in } \mathfrak{X} - E_A)$$

provided that $\mu(X) \leq \in /2A = \delta$ □

Theorem 26. *If $f(x)$ is integrable on X and $X_n \uparrow X$, then*

$$F(X_n) \to F(X).$$

Proof. If $\mu(X) < \infty$ this follows from Theorem 25 and the continuity of μ in the sense of Theorem 2. If $\mu(X) = \infty, \in > o$ we can choose a subdivision $\{ y_\nu \}$ and corresponding subsets E_ν of X so that

$$\sum_{\nu=1}^{\infty} y_\nu \mu(E_\nu) > F(X) - \in$$

(assuming that $f(x) \geq 0$, as we may)

But

$$F(X_n) = \sum_{\nu=1}^{\infty} F(X_n E_\nu)$$

and $F(X_n E_\nu) \to F(E_\nu)$ as $n \to \infty$ for every ν, since $\mu(E_\nu) < \infty$. Since all the terms $y_\nu F(X_n E_\nu)$ are positive, it follows that 34

$$\lim_{n \to \infty} F(X_n) = \sum_{\nu=1}^{\infty} \text{F}(E_\nu) \geq \sum_{\nu=1}^{\infty} y_\nu \mu(E_\nu) > \text{F(X)} - \in$$

Since $F(X_n) \leq F(X)$, the theorem follows. □

Theorem 27. *If $f(x)$ is integrable on X and $\in > 0$, we can find a subset X_1 of X so that $\mu(X_1) < \infty$, $\int_{X - X_1} | f(x) | d\mu < \in$ and $f(x)$ is bounded in X_1.*

Proof. The theorem follows at once from Theorems 25 and 26 since we can take $X_1 \subset \in [f(x) \geq y_1]$ and this set has finite measure since f(x) is integrable. □

Theorem 28. *If $f(x)$ and $g(x)$ are integrable on $\mathfrak{X}$, so is $f(x)+g(x)$ and*

$$\int_{\mathfrak{X}} [f(x)+g(x)]d\mu = \int_{\mathfrak{X}} f(x)d\mu + \int_{\mathfrak{X}} g(x)d\mu.$$

Proof. Since $| f(x)+g(x) | \leq | f(x) | + | g(x) | \leq 2 \sup(| f(x) |, | g(x) |)$ we have

$$\int_{\mathfrak{X}} | f(x)+g(x) | d\mu \leq 2 \int_{\mathfrak{X}} \sup(| f(x) |, | g(x) |) d\mu$$

$$= 2\left[\int_{|f|\geq|g|} | f(x) | d\mu + \int_{|f|<|g|} | g(x) | d\mu \leq 2 \int_{\mathfrak{X}} | f(x) | d\mu + 2 \int_{\mathfrak{X}} | g(x) | d\mu \right]$$

so that $f(x)+g(x)$ is integrable. After Theorem 27, there is no loss of generality in supposing that $\mu(\mathfrak{X}) < \infty$. Moreover, by subdividing $\mathfrak{X}$ into the sets (not more than 8) in which $f(x)$, $g(x)$, $f(x)+g(x)$ have constant signs, the theorem can be reduced to the case in which $f(x) \geq 0$, $g(x) \geq 0$ and so $f(x)+g(x) \geq 0$ in $\mathfrak{X}$. □

35 The conclusion is obvious if $f(x)$ is a constant $c \geq 0$, for we can then take as subdivisions, $\{y_\nu\}$ for $g(x)$ and $\{y_\nu + c\}$ for $g(x)+c$. In the general case, if

$$E_\nu = \varepsilon[y_\nu \leq g(x) < y_{\nu+1}]$$

$$\int_{\mathfrak{X}} [f(x)+g(x)]d\mu = \sum_{\nu=0}^{\infty} \int_{E_\nu} [f(x)+g(x)]d\mu, \text{by Theorem 20}$$

$$\geq \sum_{\nu=0}^{\infty} \int_{E_\nu} f(x)d\mu + \sum_{\nu=1}^{\infty} y_\nu \mu(E_\nu$$

$$= \int_{\mathfrak{X}} f(x)d\mu + S,$$

and since $\int_{\mathfrak{X}} g(x)d\mu$ is sup s for all subdivisions $\{y_\nu\}$, we get

$$\int_{\mathfrak{X}} [f(x) + g(x)]d\mu \geq \int_{\mathfrak{X}} f(x)d\mu + \int_{\mathfrak{X}} g(x)d\mu$$

On the other hand, if $\in > 0$, and we consider subdivisions for which

$$y_1 \leq \in, y_{\nu+1} \leq (1+\in) y_\nu \text{ for } \nu \geq, 1$$

we get

$$\begin{aligned}\int_{\mathfrak{X}} [f(x) + g(x)]d\mu &\leq \sum_{\nu=0..}^{\infty} \int_{E_\nu} f(x)d\mu + \sum_{\nu=0}^{\infty} y_{\nu+1}\mu(E_\nu)\\ &\leq \int_{\mathfrak{X}} f(x)d\mu + (1+\in)S + y_1\mu(E_o)\\ &\leq \int_{\mathfrak{X}} f(x)d\mu + (1+\in)\int_{\mathfrak{X}} g(x)d\mu + \in \mu(\mathfrak{X})\end{aligned}$$

and the conclusion follows if we let $\in \to 0$.

Combining this result with Theorem 21, we get

Theorem 29. *The integrable functions on $\mathfrak{X}$ form a linear space over R on which $\int_{\mathfrak{X}} f(x)d\mu$ is a linear functional.*

This space is denoted by $L(\mathfrak{X})$, and $f(x)\varepsilon L(\mathfrak{X})$ means that $f(x)$ is **36**
(absolutely) integrable on $\mathfrak{X}$.

13. Convergence theorems

Theorem 30 (Fatou's Lemma)**.** *If $\gamma(x)$ is integrable on $\mathfrak{X}$, and $f_n(x)$, $n = 1, \ldots$ are measurable functions, then*

$$\limsup \int_{\mathfrak{X}} f_n(x)d\mu \leq \int_{\mathfrak{X}} (\limsup f_n(x))d\mu \text{ if } f_n(x) \leq \gamma(x),$$

$$\liminf \int_{\mathfrak{X}} f_n(x)d\mu \geq \int_{\mathfrak{X}} (\liminf f_n(x))d\mu \text{ if } f_n(x) \geq -\gamma(x),$$

As immediate corollaries we have

Theorem 31 (Lebesgue's theorem on dominated convergence)**.** *If $\gamma(x)$ is integrable on $\mathfrak{X}$, $\mid f_n(x) \mid\leq \gamma(x)$ and*

$$f_n(x) \to f(x) \text{ p.p. in } \mathfrak{X}$$

then

$$\int_{\mathfrak{X}} f_n(x)d\mu \to \int_{\mathfrak{X}} f(x)d\mu$$

In particular, the conclusion holds if $\mu(\mathfrak{X}) < \infty$ and the $f_n(x)$ are uniformly bounded.

Theorem 32 (Monotone convergence theorem)**.** *If $\gamma(x)$ is integrable on $\mathfrak{X}$, $f_n(x) \geq -\gamma(x)$ and $f_n(x)$ is an increasing sequence for each x, with limit $f(x)$ then*

$$\lim_{n\to\infty} \int_{\mathfrak{X}} f_n(x)d\mu = \int_{\mathfrak{X}} f(x)d\mu$$

in the sense that if either side is finite, then so is the other and the two values are the same, and if one side is $+\infty$, so is the other.

37 *Proof of Fatou's lemma*

The two cases in the theorem are similar. It is sufficient to prove the second, and since $f_n(x) + \gamma(x) \geq 0$, there is no loss of generality in supposing that $\gamma(x) = 0$, $f_n(x) \geq 0$,

Let $f(x) = \liminf f_n(x)$ and suppose that $\int_{\mathfrak{X}} f(x)d\mu < \infty$. Then after Theorem 27, given $\in > 0$ we can define X_1 so that $\mu(X_1) < \infty$ and $\in > \int_{\mathfrak{X}-X_1} f(x)d\mu$ while $f(x)$ is bounded in X_1.

A straight-forward modification of Egoroff's theorem to gether with theorem 25 shows that we can find a set $X_2 \subset X_1$ so that

$$\int_{X_1-X_2} f(X)d\mu < \in$$

while

$$f_n(x) \geq f(x) - \in /\mu(X_1)$$

for all x in X_2 and $n \geq$ N. Then

$$\int_{\mathfrak{X}} f_n(x)d\mu \geq \int_{X_2} f_n(x)d\mu \geq \int_{X_2} f(x)d\mu - \in$$
$$\geq \int_{\mathfrak{X}} xf(x)d\mu - 3 \in \text{ for } n \geq N$$

and our conclusion follows.

If

$$\int_{\mathfrak{X}} f(x)d\mu = \infty$$

it follows from the definition of the integral that $A > 0$, we can define $\varphi(x) \in L(\mathfrak{X})$ so that

$$\int_{\mathfrak{X}} \varphi(x)d\mu \geq A, \quad 0 \leq \varphi(x) \leq f(x)$$

The argument used above now shows that **38**

$$\int_{\mathfrak{X}} f_n(x)d\mu \geq \int_{\mathfrak{X}} \varphi(x)d\mu - 3 \in \geq A - 3 \in$$

for sufficiently large n, and hence

$$\liminf \int_{\mathfrak{X}} f_n(x)d\mu = \infty.$$

Restatement of Theorems 31 and 32 in terms of series, rather than sequences given us

Theorem 33. *(Integration of series) If $u_n(x)$ is measurable for each n, $u(x) = \sum_{n=1}^{\infty} u_n(x)$, then*

$$\int_{\mathfrak{X}} u(x)d\mu = \sum_{n=1}^{\infty} \int_{\mathfrak{X}} u_n(x)d\mu$$

provided that $|\sum_{\nu=1}^{n} u_\nu(x)| \leq \gamma(x)$ *for all* N *and* x, $\gamma(x) \in L(\mathfrak{X})$.

The equation is true if $u_n(x) \geq 0$, *in the sense that if either side is finite, then so is the other and equality holds, while if either side is* ∞ *so is the other.*

Theorem 34. *(Differentiation under the integral sign)*

If $f(x,y)$ *is integrable in* $a < x < b$ *in a neighbourhood of* $y = y_\circ$ *and if* $\dfrac{\partial f}{\partial y_\circ}$ *exists in* $a < x < b$, *then*

$$\frac{d}{dy_\circ}\int_a^b f(x,y)dx = \int_a^b \frac{\partial f}{\partial y_\circ}dx$$

provided that

$$\left|\frac{f(x,y_\circ+h) - f(x,y_\circ)}{h}\right| \leq \gamma(x)\varepsilon L(a,b)$$

39 *for all sufficiently small* h.

This theorem follows from the analogue of Theorem 31 with n replaced by a continuous variable h. The proof is similar.

14. The Riemann Integral

If we proceed to define an integral as we have done, but restrict the set function to one defined only on a *finitely* additive system of sets (we call this set function "measure" even now), we get a theory, which in the case of functions of a real variable, is equivalent to that of Riemann. It is then obvious that an-R-integrable function is also L-integrable and that the two integrals have the same value.

The more direct definition of the R-integral is that $f(x)$ is R-integrable in $a \leq x \leq b$ if it is bounded and if we can define two sequences $\{\varphi_n(x)\}, \{\psi_n(x)\}$ of step functions so that $\varphi_n(x) \uparrow, \psi_n(x) \downarrow$, for each x,

$\varphi_n(x) \leq f(x) \leq \psi_n(x), \int_a^b (\psi_n(x) - \varphi_n(x))dx \to 0$ as $n \to \infty$ since lim $\varphi_n(x) = lim \Psi_n(x) = f(x)$ p.p., it is clear that $f(x)$ is L-integrable and that its L-integral satisfies

$$\int_a^b f(x)dx = \lim_{n\to\infty} \int_a^b \varphi_n(x)dx = \lim_{n\to\infty} \int_a^b \psi_n(x)dx,$$

and the common value of these is the R-integral. The following is the main theorem.

Theorem 35. *A bounded function in (a, b) is R-integrable if and only if it is continuous p.p* 40

Lemma. *If $f(x)$ is R-integrable and $\in > 0$, we can define $\delta > 0$ and a measurable set $E_\circ$ in (a, b) so that*

$$\mu(E_0) > b - a - \in,$$
$$|f(x + h) - f(x)| \leq \in \text{ for } x \in E_0, x + h \in (a, b), |\text{h}| < \delta.$$

Proof of Lemma: We can define continuous functions $\varphi(x), \psi(x)$ in $a \leq x \leq b$ so that

(i) $\varphi(x) \leq f(x) \leq \psi(x), a \leq x \leq b$

(ii) $\int_a^b (\psi(x) - \varphi(x))dx \leq \in^2 /2$

If E_0 is the set in (a, b) in which $\psi(x) - \varphi(x) < \in /2$ it is plain that $\mu(E_0) > b - a - \in$. For otherwise, the integral in(ii) would exceed $\in^2 /2$. By uniform continuity of $\varphi(x), \psi(x)$, we can define $\delta = \delta(\in) > 0$ so that

$$\psi(x + h) - \psi(x)1 \leq \epsilon/2, |\varphi(x + h) - \varphi(x)| \leq \epsilon/2$$

for x, $x + h$ in (a, b), $|h| \leq \delta$.

Then, if x is in E_0, $x + h$ is in (a, b) and $|h| \leq \delta$

$$f(x + h) - f(x) \leq \psi(x + h) - \varphi(x) = \psi(x) - \varphi(x) + \psi(x + h) - \psi(x)$$

$$\leq \epsilon/2 + \epsilon/2 = \epsilon$$

and similarly $f(x+h) - f(x) \geq -\epsilon$, as we require.

Proof of Theorem 35 41

If $f(x)$ is R-integrable, let

$$\in > 0, \in_n > 0, \sum_{n=1}^{\infty} \in_n < \in,$$

and define measurable sets E_n in (a, b) by the lemma so that

$$\mu(E_n) > b - a - \in_n, |f(x+h) - f(x)| < \in_n \text{ for } x \varepsilon E_n,$$
$$|h| \leq \delta_n, \delta_n = \delta_n(\in_n) > 0.$$

Let $E^* = \bigcap_{n=1}^{\infty} E_n$, so that

$$\mu(E^*) \geq b - a - \sum \in_n > b - a - \epsilon$$

Since $f(x)$ is continuous at every point of E^* and $\in$ is arbitrarily small, $f(x)$ is continuous p.p.

Conversely, suppose that $f(x)$ is continuous p.p. Then if $\epsilon > 0$ we can define E_0 so that

$$\mu(E_0) > b - a - \epsilon \text{ and } \delta > 0 \text{so that}$$
$$|f(x+h) - f(x)| < \in \text{ for } x \epsilon E_0, |h| < \delta$$

If now we divide (a, b) into intervals of length at most δ those which contain a point of E_0 contribute not more than $2 \in (b - a)$ to the difference between the upper and lower Riemann sums S, s for $f(x)$, while the intervals which do not contain points of E_0 have total length ϵ at most and contribute not more than $2 \in M$ where $M = \sup |f(x)|$. Hence

$$S - s \leq 2 \in (b - a) + 2 \in M$$

42 which can be made arbitrarily small.

15. Stieltjes Integrals

In the development of the Lebesgue integral, we have assumed that the measure μ is non-negative. It is easy to extend the theory to the case in which μ is the difference between two measures μ^+ and μ^- in accordance with Theorem 3.In this case, we define

$$\int_{\sqrt{x}} f(x)d\mu = \int_{\sqrt{x}} f(x)d\mu^+ - \int_{\sqrt{x}} f(x)d(-\mu^-),$$

when both integrals on the right are finite, and since μ^+ and μ^- are measure, all our theorems apply to the integrals separately and therefore to their sum with the exception of Theorems 22, 23, 24, 30, 32 in which the sign of μ obviously plays a part. The basic inequality which takes the place of Theorem 22 is

Theorem 36. *If $\mu = \mu^+ + \mu^-$ in accordance with Theorem 3, and $\mu^- = \mu^+ - \mu^-$ then*

$$|\int_{\sqrt{x}} f(x)d\mu| \leq \int_{\sqrt{x}} |f(x)|d\mu^-.$$

[The integral on the right is often written $\int_{\sqrt{x}} |f(x)|\ |d\mu|$.]

Proof.

$$\begin{aligned}
\left|\int_{\sqrt{x}} f(x)d\mu\right| &= \left|\int_{x} f(x)d\mu^+ - {}_{X}\ f(x)d(-\mu^-)\right| \\
&\leq \left|\int_{\sqrt{x}} f(x)d\mu^+\right| + \left|\int_{\sqrt{x}} f(x)d(-\mu^-)\right| \\
&\leq \int_{\sqrt{x}} |f(x)d\mu^+ + \int_{\sqrt{x}} |f(x)|d(-\mu^-) \\
&= \int_{\sqrt{x}} |f(x)|d\mu^- = \int_{\sqrt{x}} |f(x)|\ |d\mu|.
\end{aligned}$$

We shall nearly always suppose that μ is a measure with $\mu \geq 0$ but it **43** will be obvious when theorems do not depend on the sign of μ and these can be extended immediately to the general case. When we deal with *inequalities*, it is generally essential to restrict μ to the positive case (or replace it by $\overline{\mu}$). □

Integrals with μ taking positive and negative values are usually called *Stieltjes integrals*. If they are integrals of functions $f(x)$ of a real variable. x with respect to μ defined by a function $\psi(x)$ of bounded variation, we write

$$\int\limits_X f(x) d\psi(x) \text{ for } \int\limits_X f(x) d\mu,$$

and if X is an interval (a, b) with $\psi(x)$ continuous at a and at b, we write it as

$$\int\limits_a^b f(x) d\,\psi(x).$$

In particular, if $\psi(x) = x$, we get the classical Lebesgue integral, which can always be written in this from.

If $\psi(x)$ is not continuous at a or at b, the integral will generally depend on whether the interval of integration is open or closed at each end, and we have to specify the integral in one of the four forms.

$$\int\limits_{a\pm o}^{b\pm 0} f(x) d\psi\ (x)$$

Finally, if $f(x) = F_1(x) + i f_2(x), (f_1(x), f_2(x)$ real) is a complex
44 valued function, it is integrable if f_1 and f_2 are both integrable if we define

$$\int\limits_X f(x) d\mu = \int\limits_X f_1(x) d\mu + i \int\limits_X f_2(x) d\mu.$$

The inequality

$$\left| \int\limits_X f(x) d\mu \right| \leq \int\limits_X |\ f(x)\ |\ \ |\ d\mu\ |$$

(Theorem 36) still holds.

16. *L*-Spaces

A set L of elements f, g, ... is a *linear* space over the field R of real numbers (and similarly over any field) if

(1) L is an abelian group with operation denoted by $+$.

(2) $\propto f$ is defined and belongs to L for any α of R and f of L.

(3) $(\alpha+\beta)f = \alpha f + \beta f$

(4) $\alpha(f+g) = \alpha f + \alpha g$

(5) $\alpha(\beta f) = (\alpha\beta) f$

(6) $1.f = f$.

A linear space is a *topological linear space if*

(1) L is a topological group under addition,

(2) scalar multiplication by α in R is continuous in this topology. L is a *metric linear space* if its topology is defined by a metric.

It is a Banach space if

(1) L is a metric linear space in which metric is defined by $d(f,g) = ||f-g||$ where the *norm* $||f||$ is defined as a real number for all f of L and has the properties

$||f|| = 0$ if and only if $f = 0$, $||f|| \geq 0$ always **45**

$$||\alpha f|| = |\alpha| \; ||f||, \; ||f+g|| \leq ||f|| + ||g||$$

and

(2) L is *complete*. That is, if a sequence f_n has the property that $||f_n - f_m|| \to 0$ as m, $n \to \infty$, then there is a limit f in L for which $||f_n - f|| \to 0$. A Banach space L is called a *Hilbert space* if and inner product (f,g) is defined for every f, g of L as a complex number and

(1) (f, g) is a linear functional in f and in g

(2) $(f, g) = \overline{(g, f)}$

(3) $(f, f) = ||f||^2$

Two point f, g are *orthogonal* if $(f, g) = 0$, It is obvious that the integrable functions $f(x)$ in $\mathfrak{X}$ form a linear space $L(\mathfrak{X})$ on which $\int_{\mathfrak{X}} f(x)d\mu$ is a linear functional. If $p \geq 1$ the space of measurable functions $f(x)$ on $\mathfrak{x}$ for which | f(x) |^p is integrable is denoted by $L_p(\mathfrak{X})$ and we have the fallowing basic theorems.

Theorem 37. *(Holder's inequality; Schwartz' inequality if p=2)* *If* $p \geq 1, \frac{1}{p} + \frac{1}{p'} = 1$, $f(x) \in L_p(\mathfrak{X})$ *then*

$$\left| \int_{\mathfrak{X}} f(x)g(x)d\mu \right| \leq \left(\int_{\mathfrak{X}} f(x) \mid^p d\mu \right)^{1/p} \left(\int_{\mathfrak{X}} \mid g(x) \mid^{p'} d\mu \right)^{1/p'}$$

If $p = 1$, $\left(\int_{\mathfrak{X}} \mid g(x) \mid^{p'} d\mu \right)^{1/p'}$ *is interpreted as the essential upper bound of* $g(x)$ *that is, the smallest number* Λ *for which* $\mid g(x) \mid \leq \Lambda$ *p.p*

46 **Theorem 38.** *If* $q \geq p \geq 1$ *and* $\mu(\chi) < \infty$*, then*

$$L_q(\mathfrak{X}) \subset L_p(\mathfrak{X}).$$

If $\mu(\mathfrak{X}) = \infty$*, there is no inclusion relation between* L_p*,* L_q*. For the proof we merely apply Holder's theorem with* $f(x)$ *,*$g(x)$*, p replaced by* $|f(x)|^p$*, 1,* $\dfrac{q}{p}$ *respectively.*

Theorem 39 (Minkowski's Inequality)**.**
If $p \geq 1$ *and* $||f|| = (\int_x |f(x)|^p d\mu)^{1/p}$ *, then*

$$||f + g|| \leq ||f|| + ||g||$$

For the proofs see Hardy, Littlewood and Polya: Inequalities.

Theorem 40. *If* $p \geq 1, L_p(\chi)$ *is complete. (the case* $p = 2$ *is the Riesz-Fischer theorem).*

Proof. We support that $p < \infty$ and that

$$\|f_n - f_m\| \to 0 \ \ as \ \ m, n \to \infty,$$

(in the notation introduced in Theorem 39) and define $A_k > 0 \in_k \downarrow 0$ so that $\sum A_k < \infty$ and $\sum(\in_k / A_k)^p < \infty$.

We can choose a sequence $\{n_k\}$ so that $n_{k+1} > n_k$ and

$$\|f_{n_k} - f_m\| \leq \in_k \text{ for } m \geq n_k$$

and in particular

$$\|f_{n_{k+1}} - f_{n_k}\| \leq \epsilon_k.$$

Let E_k be the set in which $|f_{n_{k+1}}(x) - f_{n_k}(x)| \leq A_k$. □

Then 47

$$\in_k^p \geq \int\limits_{\mathfrak{X}} |f_{n_{k+1}}(x) - f_{n_k}(x)|^p \, d\mu \geq \int\limits_{\mathfrak{X}-E_k} |f_{n_{k+1}}(x) - f_{n_k}(x)|^p \, d\mu$$
$$\geq A_k^p \mu(\mathfrak{X} - E_k),$$

so that $\mu\left[\bigcup\limits_K^\infty (\mathfrak{X} - \in_k)\right] \to 0$ as $K \to \infty$ since $\sum(\in_k / A_k)^p < \infty$.

Since $f_{n_k}(x)$ tends to a limit at every point of each set $\bigcap\limits_K^\infty E_k$ (because $\sum A_k < \infty$) , it follows that $f_{n_k}(x)$ tends to a limit $f(x)$ p.p.

Also, it follows from Fatou's lemma that, since $\|f_{n_k}\|$ is bounded, $f(x) \in L_p(\mathfrak{X})$ and that

$$\|fn_k - f\| \geq \in_k, \|fn_k - f\| \to 0 \ \ as \ \ k \to \infty$$

Since $\|fn_k - f_m\| \to 0 \ as \ k, m \to \infty$ it follows from Minkowski's inequality that $|f_m - f\| \to 0 \ \ as \ \ m \to \infty$.

If $p = \infty$, the proof is rather simpler.

From these theorems we deduce

Theorem 41. *If $p \geq 1$, $L_p(\mathfrak{X})$ is a Banach space with*

$$\|f\| = \left(\int_{\mathfrak{X}} |f(x)|^p \ d\mu \right)^{1/p}$$

L_2 is a Hilbert space with

$$(f, g) = \int_{\mathfrak{X}} f(x) \ \overline{g(x)} \, d\mu.$$

The spaces L_p generally have certain separability properties related to the topological properties (if any) of $\mathfrak{X}$.

48 A function with real values defined on an additive system S_0, taking constant values on each of a finite number of sets of S_0 id called a *step function*.

Theorem 42. *The set of step functions (and even the sets step function taking relation values) is dense in L_p for $1 \leq p < \infty$. If the Borel system of measurable sets in $\mathfrak{X}$ is generated by a countable, finitely additive system S_0 ,then the set of steps functions with rational values is countable and L_p is separable.*

The proof follows easily from the definition of the integral.

Theorem 43. *If every step function can be approximated in $L_p(\mathfrak{X})$ by continuous functions, the continuous functions, are dense in $L_p(\mathfrak{X})$, (assuming of course that $(\mathfrak{X})$ is a topological space).*

In particular, continuous functions in R_n are dense in $L_p(R_n)$. Since the measure in R_n can be generated by a completely additive function on the finitely additive countable system of finite unions of rectangles $a_i \leq x_i < b_i$, and with a_i, b_i rational their complements, $L_p(R_n)$ is separable.

We have proves in Theorem 25 that an integral over an arbitrary set X is an absolutely continuous function of X. The following theorem provides a converse.

Theorem 44. *(Lebesgue for R_n; Radon-Nikodym in the general case)*

If $H(x)$ is completely additive and finite in $\mathfrak{X}$ and if $\mathfrak{X}$ has finite mea-
49 *sure or is the limit of a sequence of subset of finite measure, then*

$$H(X) = F(X) + Q(X)$$

where $$F(X) = \int\limits_X f(x)d\mu,\ f(x) \in L(\mathfrak{X})$$

and Q is a (singular) function with the property that there is a set $\mathfrak{X}_s$ of measure zero for which

$$0 \leq Q(X) = Q(X.\mathfrak{X}_s)$$

for all measurable X. Moreover,$F(X), Q(X)$ are unique and $f(x)$ is unique up to a set of measure zero.

In particular, if $H(X)$ is absolutely continuous, $Q(X) = 0$ and

$$H(X) = F(X) = \int\limits_X f(x)d\mu,\ f(x) \in L(\mathfrak{X})$$

In this case,$f(x)$ is called the Radon derivative of $H(X) = F(X)$

Proof. We assume that $\mu(\mathfrak{X}) < \infty$. The extension is straightforward.

By Theorem 3, we can suppose that $H(X) \geq 0$. let Θ be the class of measurable function $\theta(x)$ with the property that $\theta(x) \geq 0$,

$$\int_x \theta(x)d\mu \leq H(X)$$

for all measurable X. Then we can find a sequence $\{\theta_n(x)\}$ in Θ for which

$$\int_{\mathfrak{X}} \theta_n(x)d\mu \longrightarrow \sup_{\theta\epsilon\Theta} \int_x \theta(x)d\mu \leq H(\mathfrak{X}) < \infty$$

□

If we define $\theta'_n(x) = \sup_{k\leq n}\theta_k(x)$ and observe that 50

$$X = \bigcup_{k=1}^{n} X.\epsilon[\theta'_n(x) = \theta_k(x)]$$

we see that $\theta'_n(x)$ belongs to Θ. Since $\theta'_n(x)$ increases with n for each x, it has a limit $f(x) \geq 0$, which also belongs to Θ ,and we can write

$$F(X) = \int_X f(x) d\mu \leq H(X),\ Q(X) = H(X) - F(X) \geq 0$$

while

(1) $\int_{\mathfrak{X}} f(x) d\mu = \sup_{\theta \epsilon \Theta} \int_{\mathfrak{X}} \theta(x) d\mu < \infty$.

Now let

$$Q_n(X) = Q(X) - \frac{\mu(X)}{n}$$

and let $\mathfrak{X}_n^+$, $\mathfrak{X}_n^-$ be the sets defined by Theorem 3 for which

$$Q_n(X) \geq 0 \text{ if } X \subset \mathfrak{X}_n^+,\ Q_n(X) \leq 0 \text{ if } X \subset \mathfrak{X}_n^+$$

Then,

$$H(X) \geq F(X) + \frac{\mu(X)}{n} = \int_X (f(x) + \frac{1}{n}) \mathrm{d}\mu \text{ if } X \subset \mathfrak{X}_n^+$$

and if

$$f(x) = f(x) + \frac{1}{n} \text{ for } x \in \mathfrak{X}_n^+$$
$$f(x) = f(x) \text{ for } x \in \mathfrak{X}_n^+,$$

it follows that $f(x)$ be longs to Θ, and this contradicts(1) unless $\mu(\mathfrak{X}_n^+) = 0$. Hence $\mu(\mathfrak{X}_n^+) = 0$ and $Q(X) = 0$ if X is disjoint from

$$\mathfrak{X}_s = \bigcup_{n=1}^{\infty} \mathfrak{X}_n^+$$

which has measure zero.

51 To prove uniqueness, suppose that the decomposition can be made in two ways so that

$$H(X) = F_1(X) + Q_1(X) = F_2(X) + Q_2(X),$$

where $F_1(X), F_2(X)$ are integrals and $Q_1(X)$, $Q_2(X)$ vanish on all sets disjoint from two sets of measure zero, whose union $\mathfrak{X}_s$ also has measure zero. Then

$$F_1(X) = F_1(X - X\mathfrak{X}_S), F_2(X) = F_2(X - X\mathfrak{X}_S),$$
$$F_1(X) - F_2(X) = Q_2(X - X\mathfrak{X}_s) - Q_1(X - X\mathfrak{X}_s) = 0.$$

Theorem 45. *If $\phi(X)$ is absolutely continuous in $\mathfrak{X}$ and has Radon derivative $\varphi(X)$, with respect to a measure μin$\mathfrak{X}$,then*

$$\int_{\mathfrak{X}} f(x)d\phi = \int_{\mathfrak{X}} f(x)\varphi(x)d\mu$$

if either side exists.

Proof. We may suppose that $f(x) \geq 0, \varphi(x) \geq 0, \phi(X) \geq 0$, Suppose that

$$\int_{\mathfrak{X}} f(x)d\phi < \infty$$

Then, it follows from Theorem 27 that we may suppose that $\phi(\mathfrak{X}) < \infty$. If $\mathcal{E} > 0$, we consider subdivisions $\{y_\nu\}$ for which

$$y_1 \leq \epsilon, y_{\nu+1} \leq (1+\epsilon)y_\nu (\nu \geq 1)$$

so that

$$\begin{aligned} s = \sum_{\nu=1}^{\infty} y_\nu \phi(E_\nu) &\leq \int_{\mathfrak{x}} f(x)d\phi \\ &\leq \sum_{\nu=0}^{\infty} y_{\nu+1}\Phi(E) \\ &\leq (1+\in)s+\in \Phi(\mathfrak{X}) \end{aligned}$$

But 52

$$\int_{\mathfrak{X}} f(x)\phi(x)d\mu = \sum_{\nu=o}^{\infty}\int_{E_\nu} f(x)\phi(x)d\mu$$

by Theorem 20, and

$$y_\nu \Phi(E_\nu) \leq \int_{E_\nu} f(x)\varphi(x)d\mu \leq y_{\nu+1}\Phi(E_\nu)$$

by Theorem 22, and therefore we have also

$$s \leq \int_{\mathfrak{X}} f(x)\phi(x)d\mu \leq (1+\in)s+\in\Phi(x).$$

The conclusion follows on letting $\in\to 0$ □

Moreover the first part of this inequality holds even if $\int_{\mathfrak{X}} f(x)d\mu = \infty$, but in this case, s is not bounded and since the inequality holds for all s,

$$\int_{\mathfrak{X}} f(x)\varphi(x)d\mu = \infty.$$

17. Mappings of measures

Suppose that we have two spaces $\mathfrak{X}$, $\mathfrak{X}^*$ and a mapping $X \to X^*$ of $\mathfrak{X}$ into $\mathfrak{X}^*$. If S is a Borel system of measurable sets X with a measure μ in $\mathfrak{X}$, the mapping induces a Borel system S^* of 'measurable' sets X^* in $\mathfrak{X}^*$, these being defined as those sets X^* for which the inverse images X in $\mathfrak{X}$ are measurable, the measure μ^* induced by μ on $s*$ being defined by $\mu^*(x^*) = \mu(x)$ where X is the inverse image of X^*.

If the mapping is (1-1), the two spaces have the same properties of measure and we call the mapping a measure isomorphism.

53 **Theorem 46** (Change of variable)**.** *If the measure μ, μ^* in $\mathfrak{X}$ and $\mathfrak{X}^*$ are isomorphic under the (1-1) mapping $X \to X^*$ of $\mathfrak{X}$ onto $\mathfrak{X}^*$ and if $f^*(x^*) = f(x)$ then*

$$\int_{\mathfrak{X}} f(x)d\mu = \int_{\mathfrak{X}^*} f^*(x^*)d\mu^*$$

The proof is immediate if we note that the sets E_ν and $E^*{}_\nu$ defined in $\mathfrak{X}$ and $\mathfrak{X}^*$ respectively by any subdivision correspond under the mapping $x \to x^*$ and have the same measure $\mu\,(E_\nu) = \mu^*(E^*{}_\nu)$.

As an immediate corollary of this theorem we have

Theorem 47. *If $\alpha(t)$ increases for $A \leq t \leq b$ and $\alpha(A) = a$, $\alpha(B) = b$ and $G(x)$ is of bounded variation in $a \leq x \leq b$, then*

$$\int_a^b f(x)dG(x) = \int_A^B f(\alpha(t))dG(\alpha(t)).$$

In particular

$$\int_a^b f(x)dx = \int_A^B f(\alpha(t))d\alpha(t)$$

and, if $\alpha(t)$ is absolutely continuous

$$\int_a^b f(x)dx = \int_A^B f(\alpha(t))\alpha'(t)dt.$$

18. Differentiation

It has been shown in Theorem 44 that any completely additive and absolutely continuous finite set function can be expressed as the the integral of an integrable function defined uniquely upto a set of measure zero called its Radon derivative. This derivative does not depend upon any topological properties of the space $\mathfrak{X}$. On the other hand the derivative **54**
of a function of a real variable is defined, classically, as a limit in the topology of R. An obvious problem is to determine the relationship between Radon derivatives and those defined by other means. We consider here only the case $\mathfrak{X} = R$ where the theory is familiar (but not easy). We need some preliminary results about derivatives of a function $F(x)$ in the classical sense.

Definition. *The upper and lower, right and left derivatives of $F(x)$ at x are defined respectively, by*

$$D^+F = \lim_{h\to+0} \sup \frac{F(x+h) - F(x)}{h}$$
$$D_+F = \lim_{h\to+0} \inf \frac{F(x+h) - F(x)}{h}$$
$$D^-F = \lim_{h\to-0} \sup \frac{F(x+h) - F(x)}{h}$$

$$D_{-}F = \lim_{h\to -0}\inf \frac{F(x+h)-F(x)}{h}$$

Plainly $D_{+}F \le D^{+}F$, $D_{-}F \le D^{-}F$. If $D_{+}F = D^{+}F$ or $D_{-}F = D^{-}F$ we say that $F(x)$ is differentiable on the or on the left, respectively, and the common values are called the right or left derivatives, F'_{+}, F'_{-}. If all four derivatives are equal, we say that $F(x)$ is differentiable with derivative $F'(x)$ equal to the common value of these derivatives.

Theorem 48. *The set of points at which F'_{+} and F'_{-} both are exist but different is countable.*

55 *Proof.* It is enough to prove that the set E of points x in which $F'_{-}(x) < F'_{+}(x)$is countable. Let $r_1, r_2 \ldots$ be the sequence of all rational numbers arranged in some definite order. If $x \in E$ let $k = k(x)$ be the smallest integer for which

$$F'_{-}(x) < r_k < F'_{+}(x)$$

Now let m, n be the smallest integers for which

$$r_m < x, \frac{F(\zeta)-F(x)}{\zeta - x} < r_k \text{ for } r_m < \zeta < x,$$
$$r_n > x, \frac{F(\zeta)-F(x)}{\zeta - x} > r_k \text{ for } x < \zeta < r_n$$

□

Every x defines the triple (k, m, n) uniquely, and two numbers $x_1 < x_2$ cannot have the same triple (k, m, n) associated with them. For if they did, we should have

$$r_m < x_1 < x_2 < r_n$$

and therefore

$$\frac{F(x_1)-F(x_2)}{x_1 - x_2} < r_k \quad \text{from the inequality}$$

while
$$\frac{F(x_1)-F(x_2)}{x_1 - x_2} < r_k \quad \text{from the second}$$

and these are contradictory. Since the number of triples (k, m, n) is countable, so is E.

Theorem 49 (Vitali's covering theorem)**.** *Suppose that every point of*
56 *a bounded set E of real numbers (not necessarily measurable) is contained in an arbitrarily small closed interval with positive length and belonging to a given family V. Suppose that G is an open set containing E and that $\in > 0$.*

Then we can select a *finite* number N of mutually dis joint intervals I_n of V so that each I_n lies in G and

$$\sum_{n=1}^{N} \mu(I_n) - \in \leq \mu(E) \leq \mu(E \sum_{n=1}^{N} I_n) + \in .$$

(μ standing of course, for *outer* measure).

Proof. If $\in > 0$, it is obviously enough, after Theorem 12, to prove the theorem in the case $\mu(G) \leq \mu(E) + \in$. We may also suppose that all the intervals of V lie in G. □

We define a sequence of intervals $I_1, I_2 \ldots$ inductively as follows. I_1 is an arbitrary of V containing points of E. If $I_1, I_2 \ldots, I_n$ have been defined, let 1_n be the upper bound of lengths of all the intervals of V which contain points of E and which are disjoint from $I_1 + I_2 + \cdots + I_n$. Then, since the I_k are closed, $1_n > 0$ unless $I_1 + I_2 + \cdots + I_n \supset E$. Now define I_{n+1} so that it is an interval of the type specified above and so that $\lambda_{n+1} = \mu(I_{n+1}) > \frac{1}{2} 1_n$.

Then I_{n+1} is disjoint from $I_1 + \cdots + I_n$ and

$$S = \sum_{n=1}^{\infty} I_n \subset G.$$

Suppose now that $A = E - SE$, $\mu(A) > 0$. Let J_n be the interval with that same centre as I_n and 5 times the length of I_n. We can then choose

N so that 57

$$\sum_{n=N+1}^{\infty} \mu(J_n) = 5 \sum_{n=N+1}^{\infty} \mu(I_n) < \mu(A),$$

since $\sum\limits_{n=1}^{\infty} \mu(I_n) \leq \mu(G) \leq \mu(E)+ \in < \infty$ and $\mu(A) > 0$. It follows that

$$\mu\left(A - A \bigcup_{n=N+1}^{\infty} J_n\right) > 0$$

and that $A - A \bigcup\limits_{n=N+1}^{\infty} J_n$ contains at least one point ξ. Moreover, since ξ does not belong to the *closed* set $\sum\limits_{n=1}^{N} I_n$, we can choose from V an interval I containing ξ and such that $I.\ I_n = 0$ for $n = 1, 2 \ldots, N$. On the other hand, $I.I_n$ cannot be empty for all $n \geq N + 1$ for, if it were, we should have

$$0 < \mu(I) \geq 1_n < 2\lambda_{n+1}$$

for all $n \leq N + 1$ and this is impossible since $\lambda_n \to 0$ (for $\sum\limits_{1}^{\infty} \lambda_n = \sum\limits_{1}^{\infty} \mu(I_n) \leq \mu(G) < \infty$). We can therefore define $n_\circ \geq N + 1$ to be the smallest integer for which $I.I_{n_\circ} \neq 0$. But

$$I.I_n = 0 for n \leq n_\circ - 1$$

and it follows from the definition of 1_n that

$$0 < \lambda \leq 1_{n_\circ - 1} < 2\lambda_{n_\circ}$$

Hence I, and therefore ξ, is contained in $J_{n_\circ}$ since $J_{n_\circ}$ has five times the length of $I_{n_\circ}$ and $I.I_{n_\circ} \neq 0$.

58 This is impossible since ξ belongs to $A - A \bigcup\limits_{n=N+1}^{\infty} J_n$ and $n_\circ \geq N + 1$.

Hence we must have

$$\mu(A) = 0$$

and $\mu(ES) = \mu(E)$, $\sum\limits_{n=1}^{\infty} \mu(I_n) \leq \mu(G) \leq \mu(E)+ \in$.

We can therefore choose N as large that

$$\sum_{n=1}^{N} \mu(I_n) - \in \leq \mu(E) \leq \mu(E \sum_{n=1}^{N} I_n) + \in$$

Theorem 50. *A function $F(x)$ of bounded variation is differentiable p.p.*

Proof. It is sufficient to prove the theorem when $F(x)$ is increasing. We prove that $D^+F = D_+F$ p.p. The proof that $D^-F = D_-F$ p.p. is similar, and the conclusion then follows from Theorem 48.

The set

$$\in [D^+F > D_+F] = \bigcup_{r_1, r_2} \in [D^+F > r_1 > r_2 > D_+F]$$

Where the union is over the countable pairs of rational r_1, r_2.

Hence, if we suppose that $\mu(\in [D^+F > D_+F]) > 0$ we can find rationals r_1, r_2 such that

$$D^+F > r_1 > r_2 > D_+F$$

in a set E of positive outer measure. Then every point x of E is the left hand end point of an interval (x, η) such that

$$F(\eta) - F(x) \leq (\eta - x)r_2$$

and we may suppose that $\eta - x$ is arbitrarily small. It follows from **59**
Vitali's theorem that we can define a set K consisting of a finite number of such intervals so that

$$\mu(E.K) > \mu(K) - \in$$

While the increment $F(K)$ of $F(x)$ over the intervals satisfies

$$F(K) \leq r_2 \mu(K).$$

But every point x of EK, with the exception of the finite set of right hand end points of K, is the left hand end point of an arbitrarily small interval (x, ξ) for which

$$F(\xi) - F(x) \geq (\xi - x)r_1.$$

If we now apply Vitali's theorem to the intersection of E and the set of interior pints of K (Which has the same measure as EK), we can construct a finite set of intervals K' so that

$$K' \subset K, \mu(K') \geq \mu(EK) - \in \geq \mu(K) - 2 \in,$$

while the increment $F(K')$ of $F(x)$ over the intervals K' satisfies

$$F(K') \geq r_1 \mu(K').$$

Since $F(K') \leq F(K)$, we get

$$r_2 \mu(K) \geq r_1 \mu(K') \geq r_1(\mu(K) - 2 \in),$$

which gives a contradiction if ϵ is small enough. Hence we must have $\mu(E) = 0$ and the theorem is proved. □

Theorem 51. *If $F(x)$ increases and is bounded in $a \leq x \leq b$ and if $F'(x)$*
60 *is its derivative, then $F'(x)$ is non-negative p.p, integrable in (a, b) and satisfies*

$$\int_a^b F'(x) dx \leq F(b) - F(a)$$

Proof. Since $\frac{F(x+h)-F(x)}{h} \geq 0$ for $h \neq o$ it follows that

$$F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} \geq 0 \text{ p.p.}$$

It follows now from Fatou's lemma that if $\delta > 0$,

$$\begin{aligned}
\int_a^{b-\delta} F'(x) dx &\leq \lim_{h \to 0} \inf \int_a^{b-\delta} \frac{F(x+h) - F(x)}{h} dx \\
&= \lim_{h \to 0} \inf \left\{ \frac{1}{h} \int_{a+h}^{b+h-\delta} F(x) dx - \frac{1}{h} \int_a^b F(x) dx \right\} \\
&= \lim_{h \to 0} \inf \left\{ \frac{1}{h} \int_{b-\delta}^{b+h-\delta} F(x) dx - \frac{1}{h} \int_a^{a+h} F(x) dx \right\} \\
&\leq \lim_{h \to 0} [F(b+h-\delta) - F(a)] \\
&\leq F(b) - F(a)
\end{aligned}$$

since $F(x)$ is increasing. □

Theorem 52. *IF $f(x)$ is integrable in (a, b) and*

$$F(x) = \int_a^x f(t)dt = 0 \text{ for } a \leq x \leq b$$

then $f(x) = 0$ p.p.

(This is a refinement of corollary 1 of Theorem 24, since the condition here is lighter than the condition that the *set* function $F(X) = \int_X f(t)dt$ should vanish for for *all measurable* X.) **61**

Proof. Our hypothesis implies that $F(X) = 0$ for all open or closed intervals X and therefore, since $F(X)$ is completely additive $F(X) = 0$ for all open sets X every open set being the sum of a countable number of disjoint open intervals. But every measurable set is *t* he sum of a set of zero measure and the limit of a decreasing sequence of open sets by Theorem 12, and therefore $F(X) = 0$ for every measurable set X. The conclusion then follows from corollary 1 to Theorem 24. □

Theorem 53 (Fundamental theorem of the calculus)**.** (i) *If $F(x)$ is an absolutely continuous point function and $f(x)$ is the Radon derivative of its associated set function $F(X)$ (which is also absolutely continuous ; see page 30) then $F(x)$ is differentiable p.p and*

$$F^{'}(x) = f(x) \text{ p.p}$$

(ii) *If $f(x) \varepsilon L(a, b)$ then $F(x) = \int_a^x f(t)dt$ is absolutely continuous and $F^{'}(x) = f(x)$ p.p*

(iii) *If $F(x)$ is absolutely continuous in $a \leq x \leq b$, then $F^{'}$ is integrable and*

$$F(x) = \int_a^x F'(t)dt + F(a)$$

Proof. (i) We may suppose that $F(x)$ increases and that $F(x) \geq o$. If $A > 0$, let $f_A(x) = \min[A, f(x)], F_A(x) = \int_a^x f_A(t)dt$, where $f(x)$ is the **62**
Radon derivative of $F(X)$ and $F(x) = \int_a^x f(t)dt$. □

Then since $f_a(x)$ is bounded it follows from Fatou's lemma that

$$\begin{aligned}\int_a^x F_A'(t)dt &= \int_a^x \lim_{h\to 0} \frac{F_A(t+h)-F_A(t)}{h}dt \\ &\geq \limsup_{h\to 0} \int_a^x \frac{F_A(t+h)-F_A(t)}{h}dt \\ &= \limsup_{h\to 0} \left\{\frac{1}{h}\int_x^{x+h} F_A(t)dt - \frac{1}{h}\int_a^{a+h} F_A(t)dt\right\} \\ &= F_A(x) - F_A(a) = F_A(x)\end{aligned}$$

since $F_A(t)$ is continuous. Since $f(t) \geq f_A(t)$ it follows that $F'(t) \geq F_A'(t)$ and therefore

$$\int_a^x F'(t)dt \geq \int_a^x F_A'(t)dt \geq F_A(x)$$

This holds for all $A > 0$ and since $F_A(x) \to F(x) as A \to \infty$ by Theorem 31, we deduce that

$$\int_a^x F'(t)dt \geq F(x) = \int_a^x f(t)dt.$$

Combining this with Theorem 50 we get

$$\int_a^x (F'(t) - f(t))dt = 0$$

for all x, and the conclusion follows from Theorem 51

Parts(ii) and (iii) follow easily form(i). If we did not wish to use
63 the Radon derivative, we could prove (ii) and (iii) with the help of the deduction from Vitali's theorem that if $F(x)$ is absolutely continuous and $F'(x) = 0$ p.p then $F(x)$ is constant

Theorem 54 (Integration by parts)**.** *If $F(x), G(x)$ are of bounded variation in an open or closed interval J and*

$$F(x) = \frac{1}{2}[F(x-0)+F(x+0)], G(x) = \frac{1}{2}[G(x-0)+G(x+0)],$$

then

$$\int_j F(x)dG(x) = \int_J [F(x)G(x)] - \int_j G(x)dF(x).$$

In particular if $F(x)$, $G(x)$ are absolutely continuous then,

$$\int_a^b F(x)G''(x)dx = \int_a^b [F(x)G(x)] - \int_a^b F^{'}(x)G(x)dx$$

Proof. We may suppose that F(x), G(x) increase on the interval and are non - negative and define

$$\Delta(I) = \int_I F(x)dG(x) + \int_I G(x)dF(x) - \int_I [F(x)Gx]$$

for intervals $I \subset J$. Then $\Delta(I)$ is completely additive and we shall prove that $\Delta(I) = 0$ for all I □

Suppose first that I consist of a single point a. Then

$$\begin{aligned}\Delta(I) &= F(a)[G(a+0) - G(a-0)] + G(a)[F(a+0) - F(a-0)] \\ &\qquad - F(a+0)G(a+0) + F(a-0)G(a-0) \\ &= 0\end{aligned}$$

since $2F(a) = F(a+0) + F(a-0)$, $2G(a) = G(a+0) + G(a-0)$.

Next if I is an open interval $a < x < b$,

$$\begin{aligned}\Delta(I) &\leq F(b-0)[G(b-0) - G(a+0)]G(b-0)[F(b-0) - F(a+0)] \\ &\qquad - F(b-0)G(b-0) + F(a+0)G(a+0) \\ &= (F(b-0) - F(a+0))(G(b-0) - G(a+0)), \\ &= F(I)G(I)\end{aligned}$$

where $F(I)$, $G(I)$ are the interval functions defined by F(x),G(x), and **64**
similarly

$$\Delta(I) \geq -F(I)G(I) \text{so that } |\Delta(I)| \geq F(I)G(I)$$

Now, any interval is the sum of an open interval and one or two end points and it follows from the additivity of $\Delta(I)$, that

$$|\Delta(I)| \leq F(I)G(I).$$

for all intervals. Let $\in > 0$. Then apart from a finite number of points at which $F(x+0) - F(x-0) > \in$, and on which $\Delta = 0$, we can divide I into a finite number of disjoint intervals I_n on each of which $F(I_n) \leq \in$. Then

$$| \Delta(I) | = | \Delta\left(\sum I_n\right) | = | \sum \Delta(I_n) | \leq \sum F(I_n)G(I_n)$$
$$\leq \in \sum G(I_n) = \in G(I).$$

The conclusion follows on letting $\in \to 0$.

Theorem 55 (Second Mean Value Theorem)**.** (i) *If* $f(x) \in L(a,b)$ *and* $\varphi(x)$ *is monotonic,*

$$\int_a^b f(x)\phi(x)dx = \varphi(a+0)\int_a^\xi f(x)dx + \phi(b-0)\int_\xi^b f(x)dx$$

65 *for some* ξ *in* $a \leq \xi \leq b$.

(ii) *If* $\varphi(x) \geq 0$ *and* $\phi(x)$ *decreases in* $a \leq x \leq b$,

$$\int_a^b f(x)\varphi(x)dx = \varphi(a+0)\int_a^\xi f(x)dx$$

for some ξ, $a \leq \xi \leq b$.

Proof. Suppose that $\phi(x)$ decreases in (i), so that, if we put $F(x) = \int_a^x f(t)dt$, we have

$$\int_a^b f(x)\phi(x)dx = {}_{a+0}^{b-0}[F(x)\phi(x)] - \int_{a+0}^{b-0} F(x)d(x)$$
$$= \phi(b-0)\int_a^b f(x)dx + [\phi(a+0) - \phi(b-0)]F(\xi)$$

by Theorem 22 and the fact that $F(x)$ is continuous and attains every value between its bounds at some point ξin a $\leq \xi \leq b$. This establishes (i) and we obtain (ii) by defining $\varphi(b+0) = 0$ and writing

$$\int_a^b f(x)\phi(x)dx = \int_{a+0}^{b+0}[F(x)\phi(x)] - \int_{a+0}^{b+0} F(x)d\varphi(x)$$
$$= \varphi(a+0)F(\xi) \text{ with } a \leq \xi \leq b.$$

A refinement enables us to assert that a $< \xi < b$ in (i) and that $a < \xi \leq b$ in (ii). □

19. Product Measures and Multiple Integrals

Suppose that $\mathfrak{X}, \mathfrak{X}'$ are two spaces of points x, x'. Then the space of pairs (x, x') with x in $\mathfrak{X}$, x' in $\mathfrak{X}$ is called the *product space* of $\mathfrak{X}$ and $\mathfrak{X}'$ and is written $\mathfrak{X}x\mathfrak{X}'$.

Theorem 56. *Suppose that measures μ, μ', are defined on Borel systems S, S' of measurable sets X, X' in two spaces $\mathfrak{X}$, $\mathfrak{X}$ respectively. Then* **66**
a measure m can be defined in $\mathfrak{X}x\mathfrak{X}'$ in such a way that, if X, X' are measurable in $\mathfrak{X}$, $\mathfrak{X}'$ respectively, then XxX' is measurable in $\mathfrak{X}x\mathfrak{X}'$ and

$$m(XxX) = \mu(X).\mu'(X')$$

(The measure m is called the product measure of μ and μ'). The idea of product measures is basic in the theory of probability where it is vital to observe that the product measure is not the only measure which can be defined in $\mathfrak{X}x\mathfrak{X}'$.

Proof. We define a *rectangular* set in $\mathfrak{X}x\mathfrak{X}'$ to be any set XxX' with X in S, X' in S and we define its measure $m(XxX')$ to be $\mu(X) \cdot \mu'(X')$. (An expression of the form $0 \cdot \infty$ is taken to stand for 0). We call the sum of a finite number of rectangular sets a *figure* in $\mathfrak{X}x\mathfrak{X}'$ and define its measure to be the sum of the measure of disjoint rectangular sets which go to form it. It is easy to verify that this definition is independent of the decomposition used and that the figures and their complements form a finitely additive system on which their measure is finitely additive.

After Kolmogoroff's theorem (Theorem 7), it s sufficient to show that m is completely additive on figures. Suppose that

$$\sum_{n=1}^{\infty} X_n x X'_n = X_0 x X'_0$$

where the sets on the left are disjoint. If x is any point of X_0, let $J'_n(x)$ be the set of points x' of X'_0 for which (x, x') belongs to $X_n x X'_n$. Then $J'_n(x)$ is measurable in $\mathfrak{X}'$ for each x, it has measure $\mu'(X_{n'})$ when x is in **67**
X_n and 0 otherwise. This measure $\mu^{'}(J^{'}_n(x))$ is plainly measurable as a function of x and

$$\int_{X_o} \mu'(J'_n(x)) d\mu = \mu(X_n)\mu^{'}(X^{'}_n)$$

Moreover, $\sum_{n=1}^{N} \quad J'_n(X)$ is the set of points x of X_o for which (x, x') belongs to $\sum_{n=1}^{N} X_n x X_n$. It is measurable and

$$\int_{X_o} \mu' \left(\sum_{n=1}^{N} J_n(x) \right) d\mu = \sum_{n=1}^{N} \mu(X_n) \mu^{'}(X_n^{'}).$$

But since $X_0 x X_0^{'} \sum_{n=1}^{\infty} X_n x X_n^{'}$, it follows that

$$\lim_{N\to\infty} \sum_{n=1}^{N} J'_n(x) = X'_0 \text{ for every x of } X_0,$$

and therefore
$$\lim_{N\to\infty} \mu' \left[\sum_{n=1}^{N} J_n(x) \right] = \mu(X'_0) \text{ for every x of } X_0.$$

It follows from Theorem 32 on monotone convergence that

$$\mu(X_0)\mu'(X'_0) = \lim_{N\to\infty} \int_{X_0} \mu' \left(\sum_{n=1}^{N} J'_n(x) \right) d\mu = \sum_{n=1}^{N} \mu(X_n)\mu'(X'_N)$$

and so
$$m(X_0 x X_0^{'}) = \sum_{n=1}^{\infty} m(X_n x X_n^{'})$$

which completes the proof. □

68 **Theorem 57.** *Let $\mathfrak{X}, \mathfrak{X}''$ be two measure spaces with measures μ, μ' respectively such that $\mathfrak{X}(\mathfrak{X}')$ is the limit of a sequence $|\{X_n\}(\{X'_n\})$ of measurable sets of finite measure $\mu(X_n)(\mu'(X''_n))$. Let Y be a set in $\mathfrak{X} x \mathfrak{X}'$ measurable with respect to the product measure m defined by μ, μ'. Let $Y'(x)$ be the set of points $x' \in \mathfrak{X}'$ for which $(x, x') \in Y$. Then Y' is measurable in $\mathfrak{X}'$ for almost all $x \in \mathfrak{X}$, its measure $\mu'(Y'(x))$ is a measurable function of x and*

$$\int_{\mathfrak{X}} \mu'(Y'(x)) d\mu = m(Y).$$

Proof. We note first that the theorem is trivially true if Y is a rectangular set and follows immediately if Y is the sum of a countable number of rectangular sets. Further, it is also true for the limit of a decreasing sequence of sets of this type. In the general case, we can suppose that

$$Y \subset Q, Q - Y \subset \Gamma$$

where $m(\Gamma) = 0$ and Q, Γ are limits of decreasing sequence of sums of rectangular sets. Then, if $Q^{'}(x)$, $\Gamma^{'}(x)$ are defined in the same way as $Y^{'}(x)$ we have

$$Y'(x) \subset Q'(x), Q'(x) - Y' \subset \Gamma'(x)$$

where $\Gamma'(x)$, $Q'(x)$ are measurable for almost all x. But

$$\int_{\mathfrak{X}} \mu'(\Gamma(x))d\mu = m(\Gamma) = 0$$

so that $\mu'(\Gamma^{'}(x)) = 0$ for almost all x since $\mu^{'} \geq 0$, and this is enough to show that $Y^{'}(x)$ is measurable for almost all x and that

$$\mu'(Y'(x)) = \mu'(Q^{'}(x))\text{p.p.}$$

□

Finally, 69

$$\int_{\mathfrak{X}} \mu^{'}(Y^{'}(x))d\mu = \int_{\mathfrak{X}} \mu^{'}(Q(x))d\mu = m(Q) = m(Y).$$

Theorem 58 (Fubini's theorem)**.** *Suppose $\mathfrak{X}, \mathfrak{X}^{'}$ satisfy the hypotheses of Theorem 57. If $f(x, x^{'})$ is measurable with respect to the product measure defined by $\mu, \mu^{'}$ it is measurable in x for almost all $x^{'}$ and in $x^{'}$ for almost all x. The existence of any one of the integrals*

$$\int_{\mathfrak{X}x\mathfrak{X}'} |f(x, x')|dm, \int_{\mathfrak{X}} d\mu \int_{\mathfrak{X}'} |f(x, x)|d\mu', \int_{\mathfrak{X}'} d\mu' \int_{\mathfrak{X}} |f(x, x)|d\mu$$

implies that of the other two and the existence and equality of the integrals

$$\int_{\mathfrak{X}x\mathfrak{X}'} f(x, x)dm, \int_{\mathfrak{X}} d \int_{\mathfrak{X}'} f(x, x')d\mu', \int_{\mathfrak{X}} d\mu \int_{\mathfrak{X}'} f(x, x')d\mu$$

Proof. We may obviously suppose that $f(x, x^{'}) \geq 0$. Let $\{y_\nu\}$ be a subdivision with $E_\nu = \in [y_\nu \leq f(x, x^{'}) < y_{\nu+1}]$. The theorem holds for the function equal to y_ν in $E\nu$ for $\nu = 0, 1, 2, \ldots, N$ (N arbitrary) and zero elsewhere, by Theorem 57, and the general result follows easily from the definition of the integral. □

Chapter 2

Probability

1. Definitions

A measure μ defined in a space $\mathfrak{X}$ of points x is called a *probability measure* or a *probability distribution* if $\mu(\mathfrak{X}) = 1$. The measurable sets X are called *events* and the *probability of the event* X is real number $\mu(X)$. Two events X_1, X_2 are mutually exclusive if $X_1 \cdot X_2 = 0$. 70

The statement: *x is a random variable in $\mathfrak{X}$ with probability distribution μ* means

(i) that a probability distribution μ exists $\mathfrak{X}$,

(ii) that the expression *"the probability that x belongs to X"*, where X is a given event, will be taken to mean $\mu(X) \cdot \mu(X)$ sometimes written $P(x \varepsilon X)$.

The basic properties of the probabilities of events follow immediately. They are that these probabilities are real numbers between 0 and 1, inclusive, and that the probability that one of a finite or countable set of mutually exclusive events (X_i) should take palace i.e. the probability of the event $\cup X_i$, is equal to the sum of the probabilities of the events X_i.

If a probability measure is defined in some space, it is clearly possible to work with any isomorphic measure in another space. In practice, this can often be taken to be R_k, in which case we speak of a *random real vector* in the place of a random variable. In particular, if $k = 1$

we speak of a random of a *random real number or random real vari-*
71 *able*. The probability distribution is in this case defined by a *distribution function* $F(x)$ increasing from 0 to 1 in $-\infty < x < \infty$. For example, a probability measure in any finite or countable space is isomorphic, with a probability measure in R defined by a step function having jumps p_ν at $\nu = 0, 1, 2, \ldots$, where

$$p_\nu \geq 0, \sum p_\nu = 1.$$

Such a distribution is called a *discrete* probability distribution. If $F(x)$ is absolutely continuous $F'(x)$ is called the *probability density function (or frequency function)*.

Example 1. *Tossing a coin* The space $\mathfrak{X}$ has only two points H, T with four subsets, with probabilities given by

$$\mu(0) = 0, \mu(H) = \mu(T) = \frac{1}{2}, \mu(\mathfrak{X}) = \mu(H + T) = 1.$$

If we make H, T correspond respectively with the real numbers 0,1, we get the random real variable with distribution function

$$\begin{aligned} F(x) &= 0 (x < 0) \\ &= \frac{1}{2} (0 \leq x < 1) \\ &= 1 (1 \leq x) \end{aligned}$$

Any two real numbers a,b could be substituted for 0, 1.

Example 2. Throwing a die- The space contains six points, each with probability $1/6$ (unloaded die). The natural correspondence with R gives rise to $F(x)$ with equal jumps 1/6 at 1, 2, 3, 4, 5, 6.

2. Function of a random variable

72 Suppose that x is a random variable in $\mathfrak{X}$ and that $y = \alpha(x)$ is a function defined in $\mathfrak{X}$ and taking values in a space y. Suppose that y contains a

Borel system of measurable sets Y. Then y is called a *function* of the variable x if the set $\varepsilon[\alpha(x) \in Y]$ is measurable in $\mathfrak{X}$ for every measurable Y and if we take the measure in $\mathfrak{X}$ of this set as the probability measure of Y. Note that the mapping $x \to y$ need not be one-one.

(There is a slight ambiguity in notation as x may denote either the random variable in $\mathfrak{X}$ or a generic point of $\mathfrak{X}$. In practice, there is no difficulty in deciding which is meant.)

Example 1. x being a random real variable with distribution function $(d \cdot f)F(x)$ we compute the $d \cdot f \cdot G(y)$ of $y = x^2$

$$P(y < 0) = P(x^2 < 0) = 0 \text{ so that } G(y) = 0 \text{ for } y < 0.$$

If

$$a \geq 0, P(y \leq a) = P(x^2 \leq a) = P(0 \leq x \leq \sqrt{a}) + P(-\sqrt{a} \leq x < 0),$$
$$G(a+0) = F(\sqrt{a}+0) - F(-\sqrt{a}-0).$$

Example 2. If $F(x) = 0$ for $x < 0$ and $G(y)$ is the distribution function of $y = 1/x$, x having d.f. $F(x)$ then

$$G(y) = 0 for y < 0.$$

If $a \geq 0$.

$$G(a+0) = P(y \leq a) = P\left(\frac{1}{x} \leq a\right) = P(x \geq 1/a) = 1 - F(1/a - 0).$$

Since $G(a)$ is a d.f., $G(a+0) \to 1$ as $a \to \infty$, so that F must be continuous at 0. That is, $P(x = 0) = 0$.

3. Parameters of random variables

A *parameter* of a random variable (or its distribution) is a number associated with it. The most important parameters of real distributions are 73
the following.

(i) The *mean* or *expectation* $\overline{\alpha(x)}$ of a real valued function $\alpha(x)$ of a random real variable x is defined by

$$\overline{\alpha(x)} = E(\alpha(x)) = \int_{\mathfrak{X}} \alpha(x) d\mu$$

(ii) The standard deviation or dispersion σ of a random real number about its mean is defined by

$$\begin{aligned}\sigma^2 = E(x - \overline{x})^2 &= \int_{\mathfrak{X}} (x - \overline{x})^2 d\mu \\ &= \int_{\mathfrak{X}} x^2 d\mu - 2x \int_{\mathfrak{X}} x d\mu + \overline{x}^2 \int_{\mathfrak{X}} d\mu \\ &= E(x^2) - 2\overline{x}^2 + \overline{x}^2 = E(x^2) - (E(x))^2\end{aligned}$$

$\sigma^2$is called the *variance* of x.

(iii) The *range* is the interval (r, R), where

$$r = \sup_{F(a)=0} a, R = \inf_{F(a)=1} a$$

(iv) The *mean error* is $\int_{\mathfrak{X}} |x - \overline{x}| d\mu = E(|x - \overline{x}|)$

(v) A *median* is a real number A for which

$$F(A - 0) + F(A + 0) \leq 1.$$

(vi) The *mode* of an absolutely continuous distribution $F(x)$ is the unique maximum, when it exists, of $F'(X)$.

Some special distributions in R.

(i) The *Dinomial distribution*

$0 < p < 1, q = 1 - p, n$ is a positive integer. x can take values $\nu = 0, 1, 2, ..., n$ with probabilities

$$p_\nu = \binom{n}{\nu} p^\nu q^{n-\nu}, \sum_0^n p_\nu = 1.$$

74

Then

$$\overline{x} = E(\nu) = \sum_{\nu=0}^{n} \nu p_\nu = \text{np},$$

$$\sigma^2 = E(\nu^2) - \overline{x}^2 = \sum_{\nu=0}^{n} \nu^2 \text{p}_\nu - \text{n}^2\text{p}^2 = \text{npq}$$

$\mathcal{P}_\nu$ is the probability of ν successes out of n experiments in each of which the probability of success is p.

(ii) *The Poisson distribution* x can take the values $\nu = 0, 1, 2, \ldots$ with probabilities

$$p_\nu = e^{-c}\frac{c^\nu}{\nu!}, \sum_{\nu=0}^{\infty} p_\nu = 1,$$

where $c > 0$. Here

$$\overline{x} = e^{-c}\sum_{\nu=0}^{\infty}\frac{\nu c^\nu}{\nu!} = c$$

$$\sigma^2 = e^{-c}\sum_{\nu=0}^{\infty}\frac{\nu^2 c^\nu}{\nu!} - c^2 = c$$

The binomial and Poisson distributions are discrete. The Poisson distribution is the limit of the binomial distribution as $n \to \infty'$ if we put

$$p = c/n (p \text{ then } \to 0).$$

(iii) *The rectangular distribution*

This is given by

$$\begin{aligned} F(x) &= 0 \text{ for } x \leq a \\ &= \frac{x-a}{b-a} \text{ for } a \leq x \leq b \end{aligned}$$

$$= 1 \text{ for } b \leq x.$$

It is absolutely continuous and $F'(x) = 1/(b-a)$ for $a < x < b$ and = 0 for $x < a, x > b$. Also 75

$$\overline{x} = \frac{a+b}{2}, \sigma^2 = \frac{(b-a)^2}{12}$$

(iv) *The normal (or Gaussian) distribution*

This is an absolutely continuous distribution F for which

$$F'(x) = \frac{1}{\sigma\sqrt{2\prod}} \quad e^{-(x-\overline{x})^2/2\sigma^2}$$

It is easy to verify that the mean and standard deviation are respectively $\overline{x}$ and σ.

(v) *The singular distribution*:

Here $x = 0$ has probability 1,

$$F(x) = D(x-a) = \begin{array}{l} 1 \text{ if } x \geq a \\ 0 \text{ if } x < a \end{array}$$

We now prove

Theorem 1 (Tehebycheff's inequality)**.** *If* $\propto (x)$ *is a nonnegative function of a random variable* x *and* $k > 0$ *then*

$$P(\alpha(x) \geq k) \leq \frac{E(\alpha(x))}{k}$$

Proof.

$$\begin{aligned} E(\alpha(x)) &= \int_{\mathfrak{X}} \alpha(x) d\mu \\ &= \int_{\alpha(x)\geq k} \alpha(x) d\mu + \int_{\alpha(x)<k} \alpha(x) d\mu \\ &\geq \int_{\alpha(x)\geq k} \alpha(x) d\mu \geq k \int_{\alpha(x)\geq k} d\mu = KF(\alpha(x) \geq k). \end{aligned}$$

□

76

Corollary. *If $k > 0$ and $\overline{x}$, σ are respectively the mean and the standard deviation of a random real x, then*

$$p(\alpha(x) \geq k\sigma^-) \leq 1/k^2.$$

We merely replace k by $k^2\sigma^2, \alpha(x)$ by $(x - \overline{x})^2$ in the Theorem 1.

4. Joint probabilities and independence

Suppose that $\mathfrak{X}_1$, $\mathfrak{X}_2$ are two spaces of points x_1, x_2 and that a probability measure μ is define in a Borel system of sets X in their product space $\mathfrak{X}_1 x \mathfrak{X}_2$. Then the set x_1 in $\mathfrak{X}_1$ for which the set $\in [x_1 \in X_1, x_2 \in \mathfrak{X}_2]$ is measurable with respect to μ form a Borel system in $\mathfrak{X}_1$. if we define

$$\mu_1(X_1) = \mu\,(\in [x_1 \in X_1, x_2 \in \mathfrak{X}_2])$$

it follows that μ_1 is a probability measure in $\mathfrak{X}_1$ and we define μ_2 *in* $\mathfrak{X}_2$ in the same way. We call $\mu_1(X_1)$ simply the probability that x_1 belongs to X_1 with respect to the *joint distribution defined by μ.*

Definition. *If μ is the product measure of μ_1, μ_2 the random variables x_1, x_2 are said to be independent. Otherwise, they are dependant.*

When we wish to deal at the same time with several random variables, we must know their *joint probability distribution* and this, as we see that once, is not necessarily the same as the product probability as 77
their separate distributions. This applies in particular $\alpha(x_1, x_2 \ldots)$ for the probability distribution of the values of the function is determined by the joint distribution of $(x_1, x_2, \ldots)$. In this way we can define the *sum* $X_1 + X_2$ and *product* $x_1 \cdot x_2$ of random variables, each being treated as a function $\alpha(x_1, x_2)$ over the product space $\mathfrak{X}_1 x \mathfrak{X}_2$ with an assigned joint probability distribution.

Theorem 2. *If $(x_1, x_2, ..., x_n)$ is a random real vector then*

$$E(x_1 + x_2 + ... + x_n) = E(x_1) + E(x_2) + ... + E(x_n)$$

whether $x_1, x_2, \ldots, x_n$ are independent or not. ($E(x_i)$ is the mean of x_i over the product space.)

Proof. Let p be the joint probability distribution. Then

$$E\left(\sum x_i\right) = \int_{\Omega}\left(\sum x_i\right) dp \text{ (where } \Omega = \mathfrak{X}_1 x \ldots x\mathfrak{X}_n)$$

$$\sum \int_{\Omega} x_i \;\; dp = \;\; \sum E(x_i). \qquad \square$$

Theorem 3. *If $x_1, x_2, \ldots, x_n$ are independent random real variables with standard deviations $\sigma_1, \sigma_2, \ldots, \sigma_n$, then the standard deviation σ of their sum $x_i + x_2 \ldots + x_n$is given by*

$$\sigma^2 = \sum_{i=1}^{n} \sigma_i \;\; ^2$$

Proof. It is sufficient to prove the theorem for $n = 2$. Then

$$\begin{aligned}
\sigma^2 &= E((x_1 + x_2 - \overline{x}_1 - \overline{x}_2)^2) \\
&= E\left((x_1 - \overline{x}_1)^2 + (x_2 - \overline{x}_2)^2 + 2(x_1 - \overline{x}_1)(x_2\overline{x}_2)\right) \\
&= E((x_1 - \overline{x}_1)^2) + E((x_2 - \overline{x}_2)^2)2E((x_1 - \overline{x}_2)(x_2 - \overline{x}_2)) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(by theorem 2)} \\
&= \sigma_1^2 + \sigma_2^2 + 2\int\limits_{\mathfrak{X}_1 x \mathfrak{X}_2} (x_1 - \overline{x}_1)(x_2 - \overline{x}_2)dp \\
&= \sigma_1^2 + \sigma_2^2 + 2\int\limits_{\mathfrak{X}_1 (x \mathfrak{X}_2} (x_1 - \overline{x}_1)(x_2 - \overline{x}_2)d\mu_1 d\mu_2 \\
&= \sigma_1^2 + \sigma_2^2 + 2\int\limits_{\mathfrak{X}_1}(x_1 - \overline{x}_1)d\mu_1 \int\limits_{\mathfrak{X}_2}(x_2 - \overline{x}_2)d\mu_2 \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{by Fubini's theorem,} \\
&= \sigma_1^2 + \sigma_2^2
\end{aligned}$$

78 this is an example of more general principle. $\square$

Theorem 4. *If $\alpha(x_1, x_2)$ is function of two independent variables x_1, x_2 then*

$$\int_{\Omega} \alpha(x_1, x_2)dp = \iint\limits_{\Omega} \alpha(x_1, x_2)d\mu_1 \; d\mu_2, \Omega = \mathfrak{X}_1 x \mathfrak{X}_2.$$

The proof is immediate from the definition of independence. In particulars,

Theorem 5. *If x_1, x_2 are independent, then*

$$E(x_1, x_2) = E(x_1)E(x_2) = \overline{x_1}.\overline{x_2}$$

It is not generally sufficient to know the mean or other parameters of a function of random variables. The general problem is to find its complete distribution. This can be difficult, but the most important case is fairly easy.

Theorem 6. *If x_1, x_2 are independent random real numbers, with distribution functions F_1, F_2 then their sum has distribution function $F(x)$ defined by* 79

$$F(x) = F_1 * F_2(x) = F_2 * F_1(x) = \int_{-\infty}^{\infty} F_1(x-u)dF_2(u)$$

($F(x)$ is called the convolution of $F_1(x)$ and $F_2(x)$).

Proof. Let

$$\begin{aligned} \alpha_x(x_1, x_2) &= 1 \text{ when } x_1 + x_2 \leq x \\ &= 0 \text{ when } x_1 + x_2 > x \end{aligned}$$

so that if we suppose that $F(x+0) = F(x)$ and put $\Omega = R\text{x}R$, we have

$$\begin{aligned} F(x) &= \iint_\Omega \alpha_x(x_1, x_2)dp \\ &= \iint_\Omega \alpha_x(x_1, x_2)dF_1(x_1)dF_2(x_2) \\ &= \int_{-\infty}^{\infty} dF_2(x_2) \int_{-\infty}^{\infty} \alpha_x(x_1, x_2)dF_1(x_1) \\ &= \int_{-\infty}^{\infty} dF_2(x_2) \int_{x_1+x_2 \leq x} dF_1(x_1) \end{aligned}$$

$$= \int_{-\infty}^{\infty} F_1(x - x_2) dF_2(x_2) = \int_{-\infty}^{\infty} F_1(x - u) dF_2(u),$$

and a similar argument shows that

$$F(x) = \int_{-\infty}^{\infty} F_2(x - u) dF_1(u).$$

□

80 It is obvious that $F(x)$ increases, $F(-\infty) = 0$, $F(+\infty) = 1$ so that $F(x)$ is a distribution function. Moreover, the process can be repeated any finite number of times and we have

Theorem 7. *If $x_1, \ldots, x_n$ are independent, with distribution functions $F_1, \ldots, F_n$, then the distribution function of $x_1 + \ldots + x_n$ is*

$$F_{1^*} \ldots * F_n$$

Corollary . *The convolution operator applied to two or more distribution functions (more generally, functions of bounded variation in (- co, co)) is commutative and associative.*

Theorem 8. *If $F_1(x), F_2(x)$ are distribution functions, and $F_1(x)$ is absolutely continuous with derivative $f_1(x)$ then F(x) is absolutely continuous and*

$$f(x) = F'(x) = \int_{-\infty}^{\infty} f_1(x - u) dF_2(u) p.p$$

If both $F_1(x)$ and $F_2(x)$ are absolutely continuous, then

$$f(x) = \int_{-\infty}^{\infty} f_1(x - u) f_2(u) du \; p.p$$

Proof. We write

$$F(x) = \int_{-\infty}^{\infty} F_1(x - u) dF_2(u) = \int_{-\infty}^{\infty} dF_2(u) \int_{-\infty}^{x-u} f_1(t) dt$$

by the fundamental theorem of of calculus

$$= \int_{-\infty}^{\infty} dF_2(u) \int_{-\infty}^{x} f_1(t-u)dt$$
$$= \int_{-\infty}^{x} dt \int_{-\infty}^{\infty} f_1(t-u)dF_2(u)$$

and so

$$f(x) = F'(x) = \int_{-\infty}^{\infty} f_1(x-u)dF_2(u) p.p$$

again by the fundamental theorem of the calculus. The second part 81
follows from Theorem 45, Chapter I,

We shall need the following general theorem on convolutions. □

Theorem 9. *Suppose that $F_1(x)$, $F_2(x)$ are distribution functions and that $\alpha(x)$ is bounded and is either continuous or is the limit of continuous functions. Then*

$$\int_{-\infty}^{\infty} \alpha(x+y)dF_2(x) \textit{ is } B \textit{ - measurable as a}$$

function of y and

$$\int_{-\infty}^{\infty} dF_1(y) \int_{-\infty}^{\infty} \alpha(x+y)dF_2(x) = \int_{-\infty}^{\infty} \alpha(x)dF(x)$$

where

$$F(x) = F_1 * F_2(x)$$

Proof. We may suppose that $\alpha(x) \geq 0$. If we consider first the case in which $\alpha(x) = 1 fora \leq x \leq b$ and $\alpha(x) = 0$ elsewhere,

$$\int_{-\infty}^{\infty} \alpha(x+y)dF_2(x) = F_2(b-y-0) - F_2(a-y-0),$$

$$\int_{-\infty}^{\infty} dF_1(y) \int_{-\infty}^{\infty} \alpha(x+y) dF_2(x) = \int_{-\infty}^{\infty} (F_2(b-y-0)) - F_2(a-y-0) dF_1(y)$$
$$= F(b-0) - F(a-0)$$
$$= \int_{-\infty}^{\infty} \alpha(\mathrm{x}) \mathrm{dF}(\mathrm{x}),$$

82 and the theorem is true for function $\alpha(x)$ of this type.

Since an open set is the union of a countable number of intervals $a \leq x < b$, the theorem is true also for functions $\alpha(x)$ constant in each interval of an open set and 0 elsewhere. The extension to continuous function $\alpha(x)$ and their limits is immediate. □

5. Characteristic Functions

A basic tool in modern probability theory is the notion of the characteristic function $\varphi(t)$ of a distribution function $F(x)$.

It is defined by

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

Since $| e^{itx} |= 1$, the integral converges absolutely and defines $\varphi\ (t)$ for all real t.

Theorem 10. $| \varphi(t) | \leq 1$, $\varphi(0) = 1$, $\varphi(-t) = \overline{\varphi(t)}$ *and* $\varphi(t)$ *is uniformly continuous for all* t.

Proof.

$$| \varphi(t) | = \left| \int_{-\infty}^{\infty} e^{\mathrm{itx}} dF(x) \right| \leq \int_{-\infty}^{\infty} | e^{\mathrm{itx}} | \, dF(x)$$
$$= \int_{-\infty}^{\infty} dF(x) = 1.$$
$$\varphi(0) = \int_{-\infty}^{\infty} dF(x) = 1.$$

$$\varphi(-t) = \int_{-\infty}^{\infty} e^{-itx} dF(x) = \int_{-\infty}^{\infty} e^{itx} dF(x) = \overline{\varphi(t)}$$

since $F(x)$ is real. □ **83**

If h ≠ 0,

$$\varphi(t+h) - \varphi(t) = \int_{-\infty}^{\infty} e^{itx}(e^{ixh} - 1) dF(x),$$

$$| \varphi(t+h) - \varphi(t) | \leq \int_{-\infty}^{\infty} | e^{ixh} - 1 | \, dF(x) = o(1) \text{ } as \text{ } h \to 0$$

by Lebesgue's theorem, since $| e^{itx} - 1 | \leq 2$.

Theorem 11. *If $\varphi_1(t), \varphi_2(t)$ are the characteristic functions of $F_1(x)$, $F_2(x)$ respectively, then $\varphi_1(t)$. $\varphi_2(t)$ is the characteristic function of $F_1 * F_2(x)$.*

Proof.

$$\begin{aligned}
\varphi_1(t).\varphi_2(t) &= \int_{-\infty}^{\infty} e^{ith} dF_1(y). \int_{-\infty}^{\infty} e^{itx} dF_2(x) \\
&= \int_{-\infty}^{\infty} dF_1(y) \int_{-\infty}^{\infty} e^{it(x+y)} dF_2(x) \\
&= \int_{-\infty}^{\infty} dF_1(y) \int_{-\infty}^{\infty} e^{itx} dF_2(x-y) \\
&= \int_{-\infty}^{\infty} e^{itx} dF(x)
\end{aligned}$$

where $F(x) = F_1 * F_2(x)m$ by Theorem 9. □

As an immediate corollary of this and theorem 7, we have

Theorem 12. *If $x_1, x_2, \ldots, x_n$ are independent random real variables*
84 *with characteristic function $\varphi_1(t), \varphi_2(t), \ldots, \varphi_n(t)$ then the characteristic function of $x_1 + x_2 + \cdots + x_n$ is $\varphi_1(t), \varphi_2(t), \cdots, \varphi_n(t)$.*

Theorem 13. *Suppose that $F_1(x), F_2(x)$ are distribution functions with characteristic functions $\varphi_1(t), \varphi_2(t)$.*

Then

$$\int_{-\infty}^{\infty} \varphi_1(t+u)\,dF_2(u) = \int_{-\infty}^{\infty} e^{itx}\varphi_2(x)\,dF_1(x)$$

Proof.

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{itx}\varphi_2(x)\,dF_1(x) &= \int_{-\infty}^{\infty} e^{itx}\,dF_1(x) \int_{-\infty}^{\infty} e^{ixu}\,dF_2(u)\\
&= \int_{-\infty}^{\infty} dF_1(x) \int_{-\infty}^{\infty} e^{ix(t+u)}\,dF_2(u)\\
&= \int_{-\infty}^{\infty} dF_2(u) \int_{-\infty}^{\infty} e^{ix(t+u)}\,dF_1(x)\\
&= \int_{-\infty}^{\infty} \varphi_1(t+u)\,dF_2(u)
\end{aligned}$$

□

Theorem 14 (Inversion Formula)**.** *If*

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx}\,dF(x), \quad \int_{-\infty}^{\infty} |\, dF(x) \,| < \infty$$

then

(i) $F(a+h) - F(a-h) = \lim\limits_{A\to\infty} \frac{1}{\pi} \int_{-A}^{A} \frac{\sin\, ht}{t} e^{-iat}\varphi(t)\,dt$ *if $F(x)$ is continuous at $a \pm h$.*

(ii) $\int\limits_{a}^{a+H} F(x)dx - \int\limits_{a-H}^{a} F(x)dx = \frac{1}{\pi} \int\limits_{-\infty}^{\infty} \frac{1-cosHt}{t^2} e^{-iat}\varphi(t)dt.$

85

Corollary . *The expression of a function $\varphi(t)$ as an absolutely convergent Fourier - Stieltjes integral in unique. In particular, a distribution function is defined uniquely by its characteristic function.*

Proof.

$$\begin{aligned}\frac{1}{\pi}\int\limits_{-A}^{A} \cdot\frac{\sin\ \text{ht}}{t}e^{-iat}\ \varphi(t)\text{dt} &= \frac{1}{\pi}\int\limits_{-A}^{A}\frac{\sin\text{ht}}{t}e^{-iat}\text{dt}\int\limits_{-\infty}^{\infty} e^{itx}\text{dF}(x)\\ &= \frac{1}{\pi}\int\limits_{-\infty}^{\infty}\text{dF(x)}\int\limits_{-A}^{A}\frac{\sin\text{ht}}{t}e^{it(X-a)}\text{dt}\\ &= \frac{2}{\pi}\int_{-\infty}^{\infty}\text{dF(x)}\int\limits_{o}^{A}\frac{\sin\text{ht}\cos((x-a)t)}{t}\text{dt}\end{aligned}$$

But

$$\begin{aligned}= \frac{2}{\pi}\int\limits_{0}^{\infty}\frac{\sin\text{ht}\cos(x-a)t}{t}dt &= \frac{1}{\pi}\int\limits_{0}^{\infty}\frac{\sin(x-a+h)t}{t}dt\\ &\quad - \frac{1}{\pi}\int\limits_{0}^{A}\frac{\sin(x-a-h)t}{t}dt\end{aligned}$$

$$= \frac{1}{\pi}\int\limits_{0}^{A(x-a+h)}\frac{\sin t}{t}dt = \frac{1}{\pi}\int\limits_{0}^{A(x-a+h)}\frac{\sin t}{t}dt - \frac{1}{\pi}\int\limits_{A(x-ah)}^{A(x-a+h)}\frac{\sin t}{t}dt$$

and 86

$$\frac{1}{\pi}\int_{0}^{T}\frac{\sin t}{t}dt \to \frac{1}{2}, \frac{1}{\pi}\int_{-T}^{0}\frac{\sin t}{t}\,dt \to \frac{1}{2} \text{ as } T \to \infty.$$

It follows that

$$\text{A}\lim_{\to\infty}\frac{1}{\pi}\int_{A(x-a-h)}^{A(x-a+h)}\frac{\sin t}{t}dt=\begin{cases}0 & \text{if } x>a+h\\ 1 & \text{if } a-h<x<a+h\\ 0 & \text{if } x<a-h\end{cases}$$

and since this integral is bounded, it follows from the Lebesgue convergence theorem that

$$A\lim_{\to\infty}\frac{1}{\pi}\int_{-A}^{A}\frac{\sin\ ht}{t}e^{-iat}\varphi(t)dt$$
$$=\int_{a-h}^{a+h}dF(x)=F(a+h)-F(a-h)$$

provided that $F(x)$ is continuous at a $\pm h$. □

Since the integral on the left is bounded in $|\ h\ |\leq H$, we can apply Lebesgue's theorem to its integral with respect to h over $|\ h\ |\leq H$, and (ii) follows.

6. Sequences and limits of distribution and characteristic functions

Theorem 15. *If $\varphi(t)$, $\varphi_n(t)$ are the characteristic functions of distribution functions $F(x)$ and $F_n(x)$, and if $F_n(x)\to F(x)$ at every point of continuity of $F(x)$, then $\varphi_n(t)\to\varphi(t)$ uniformly in any finite interval.*

87 *Proof.* Let $\in>0$ and choose X, N $(\in)$ so that $\pm X$ are points of continuity of $F(x)$ while

$$\left(\int_{-\infty}^{-X}+\int_{X}^{\infty}\right)dF(x)<\in/2,\left(\int_{-\infty}^{-X}+\int_{X}^{\infty}dF_n(X)\right)<\in/2 \text{ for } n\geq N.$$

This is possible since the first inequality is clearly satisfied for large X and

$$\left(\int_{-\infty}^{-X}+\int_{X}^{\infty}\right)dF(X)=F_n(-X-0)+1-F_n(X+0).$$

Since F_n is a distribution function and as $n \to \infty$ this

$$\to F(-X) + 1 - F(X) = \left(\int_{-\infty}^{-X} + \int_{X}^{\infty}\right) dF(X) <\in /2$$

Since $F(x)$ is continuous at $\pm X$.

Then

$$
\begin{aligned}
| \varphi_n(t) - \varphi(t) | &\leq \in + \left| \int_{-X}^{X} e^{itx} d(F_n(X) - F(X)) \right| \\
&= \in + \left| \int_{-X}^{X} \left[e^{itx}(F_n(x) - F(x)) \right] - \int_{-X}^{X} ite^{itx}(F_n(x) - F(x)) dx \right| \\
&\leq \in + | F_n(X - 0) - F(X - 0) | + | F_n(-X + 0) - F(-X + 0) | \\
&\quad + | t | \int_{-X}^{X} | F_n(x) - F(x) | \, dx \leq \in \quad + 0(1) \text{ as } n \to \infty
\end{aligned}
$$

uniformly in any finite interval of values t, by Lebesgue's theorem. □

The converse theorem is much deeper.

Theorem 16. *If $\varphi_n(t)$ is the characteristic function of the distribution function $F_n(x)$ for $n = 1, 2, \ldots$ and $\varphi_n(t) \to \varphi(t)$ for all t, where $\varphi(t)$ is continuous at 0, then $\varphi(t)$ is continuous at 0, then φ is the characteristic function of a distribution function $F(x)$ and* **88**

$$F_n(x) \to F(x)$$

at every continuity point of $F(x)$.

We need the following

Lemma 1. *An infinite sequence of distribution functions $F_n(x)$ contains a subsequence $F_{n_k}(x)$ tending to a non-decreasing limit function $F(x)$ at every continuity point of $F(x)$. Also*

$$0 \leq F(x) \leq 1.$$

(but $F(x)$ is not necessarily a distribution function).

Proof. Let $\{r_m\}$ be the set of rational numbers arranged in a sequence. Then the numbers $F_n(r_1)$ are bounded and we can select a sequence $n_{11}, n_{12}, \ldots$ so that $F_{n_{1\nu}}(r_1)$ tends to a limit as $\gamma \to \infty$ which we denote by $F(r_1)$. The sequence $(n_{1\nu})$ then contains a subsequence $(n_{2\nu})$ so that $F_{n_{2\nu}}(r_2) \to F(r_2)$ and we define by induction sequences $(n_{k\nu}), (n_{k+1,\nu})$ being a subsequence of $(n_{k\nu})$ so that

$$F_{n_{k\nu}}(r_k) \to F(r_k) \text{as} \nu \to \infty$$

If we then define $n_k = n_{kk}$, it follows that

$$F_{n_k}(r_m) \to F(r_m) \text{ for all } m.$$

89 Also, $F(x)$ is non-decreasing on the rationals and it can be defined elsewhere to be right continuous and non-decreasing on the reals. The conclusion follows since $F(x)$, $F_{n_k}(x)$ are non-decreasing and every x is a limit of rationals. □

Proof of the theorem: We use the lemma to define a bounded non-decreasing function $F(x)$ and a sequence (n_k) so that $F_{n_k}(x) \to F(x)$ at every continuity point of $F(x)$.

If we put $a = 0$ in Theorem 14 (ii), we have

$$\int_0^H F_{n_k}(x)dx - \int_{-H}^{\circ} F_{n_k}(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1-\cos Ht}{t^2}\varphi_{n_k}(t)\mathrm{dt}$$

and if we let $k \to \infty$ and note tat $\frac{1-\cos Ht}{t^2} \varepsilon L(-\infty, \infty)$ and that F_{n_k}, $\varphi_{n_k}(t)$are bounded, we get

$$\frac{1}{H}\int_{\circ}^H F(x)dx - \frac{1}{H}\int_{-H}^{\circ} F(x)dx = \frac{1}{\pi H}\int_{-\infty}^{\infty} \frac{1-\cos Ht}{t^2}\varphi(t)\mathrm{dt}$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1-\cos t}{t^2}\varphi(\frac{t}{H})dt.$$

Now, since $\varphi(t)$ is bounded in $(-\infty, \infty)$ and continuous at 0, the expression on the right tends to $\varphi(0) = \lim_{k\to\infty} \varphi_{n_k}(0) = 1$ as

H $\to \infty$. Since F(t) is non-decreasing, it easy to show that the left hand side tends to $F(\infty) - F(-\infty)$ and hence we have

$$F(\infty) - F(-\infty) = 1,$$

and $F(x)$ is a distribution function. It now follows from Theorem 15 that φ is the characteristic function of $F(x)$.

Finally, unless $F_n(x) \to F(x)$ through the entire sequence we can define another subsequence (n_k^*) so that $F_{n_k}^*(x) \to F^*(x)$ and the same argument shows that $F^*(x)$ is a distribution function and that 90

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} dF^*(x)$$

By the corollary to Theorem 13, $F(x) = F^*(x)$, and it follows therefore that $F_n(x) \to F(x)$ at every continuity point of $F(x)$.

7. Examples of characteristic functions

Theorem 17. (i) *The binomial distribution* $p_\nu = \binom{n}{\nu} p^\nu p^{n-\nu}, \nu = 0, 1, 2, \ldots$ *has the distribution function*

$$F(x) = \sum_{\nu \leq x} p_\nu$$

and the characteristic function

$$\varphi(t) = (q + pe^{it})^n$$

(ii) *The Poisson distribution* $p_\nu = e^{-c} \frac{c^\nu}{\nu!}, \nu = 0, 1, 2, \ldots$ *has distribution function*

$$F(x) = \sum_{\nu \leq x} p_\nu$$

and characteristic function

$$\varphi(t) = e^{c(e^{it}-1)}$$

(iii) *The rectangular distribution* $F'(x) = \frac{1}{b-a}$ *for* $a < x < b$, *0 for* $x < a$, $x > b$, *0 for* $x < a$, $x > b$, *has the characteristic function*

$$\varphi(t) = \frac{e^{itb} - e^{ita}}{(b-a)it}$$

91 (iv) *The normal distribution* $F'(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$ *has characteristic function*

$$\varphi(t) = e^{-t^2\sigma^2/2}$$

(v) *The singular distribution*

$$\begin{aligned} F(x) = D(x-a) &= 0, x < a \\ &= 1, x \geq a \end{aligned}$$

has characteristic function

$$\varphi(t) = e^{ita}$$

If $a = 0$, $\varphi(t) = 1$.

These are all trivial except (iv) which involves a simple contour integration.

As a corollary we have the

Theorem 18. (i) *The sum of independent variables with binomial distributions* $(p, n_1), (p, n_2)$ *is binomial with parameters* $(p, n_1 + n_2)$.

(ii) *The sum of independent variables with Poisson distributions* c_1, c_2 *is Poisson and has parameter* $c_1 + c_2$.

(iii) *The sum of independent variables with normal distributions* $(\overline{x_1}, \delta)$ $(\overline{x_2}, \delta_2)$ *has normal distribution* $(\overline{x_1} + \overline{x_2}, \sigma), \sigma^2 = \sigma_1^2 + \sigma_2^2$.

We have also the following trivial formal result.

92 **Theorem 19.** *If x is a random real number with characteristic function* $\varphi(t)$, *distribution function* $F(x)$, *and if* A, B *are constants, then* $Ax+B$ *has distribution function* $F(\frac{X-B}{A})$ *if* $A > O$ *and* $1 - F(\frac{X-B}{A} + 0)$ *if* $A < 0$, *and characteristic function* $e^{itB}\varphi(At)$. *In particular* $-x$ *has the characteristic function* $\varphi(-t) = \overline{\varphi(t)}$.

Corollary. *If $\varphi(t)$ is a characteristic function, so is*

$$|\varphi(t)|^2 = \varphi(t)\varphi(-t).$$

The converse of Theorem 18, (ii) and (iii) is deeper. We sate it without proof. For the proof reference may be made to "Probability Theory" by M. Loose, pages 213-14 and 272-274.

Theorem 20. *If the sum of two independent real variables is normal (Poisson), then so is each variable separately.*

8. Conditional probabilities

If x is a random variable in $\mathfrak{X}$ and C is a subset of $\mathfrak{X}$ with positive measure, we define

$$P(X/C) = \mu(XC)/\mu(C)$$

to be the *conditional probability that x lies in X, subject to the condition that x belongs to C*. It is clear that $P(X/C)$ is probability measure over all measurable X.

Theorem 21 (Bayes' Theorem)**.** *Suppose that* $\mathfrak{X} = \sum_{j=1}^{J} c_j, \mu(c_j) > 0.$ *Then*

$$P(c_J/X) = \frac{P(X/C_j)\mu(C_j)}{\sum_{j=1}^{J} P(X/C_i)\mu(C_i)}$$

The proof follows at once from the definition. In applications, the sets C_j are regarded as hypotheses, the numbers $\mu(c_j)$ being called the prior probabilities. The numbers $P(C_j/X)$ are called their post probabilities or likelihoods under the observation of the event X. **93**

Example 3. Two boxes A, B are offered at random with (prior) probabilities 1/3, 2/3. A contains 8 white counters, 12 red counters, B contains 4 white and 4 red counters. A counter is taken at random from a box offered. If it turns out to be white, what is the likelihood that the box offered was A ?

If we denote the event of taking a red (white) counter by $R(W)$ the space $\mathfrak{X}$ under consideration has four points (A, R), (A, W), (B, R), (B, W). The required likelihood is.

$$P(A/W) \;=\; \frac{P(W/A)\ \mu\ (A)}{P(W/A)\ \mu\ (A) + P(W/B)\mu(B)}$$

Here

$$\begin{aligned} P(W/A) &= \text{probability of taking white counter from A} \\ &= 8/20 = 2/5 \\ P(W/B) &= 4/8 = 1/2 \end{aligned}$$

Hence

$$P(A/W) = \frac{\frac{2}{5}\frac{1}{3}}{\frac{2}{5}\frac{1}{3} + \frac{1}{2}\frac{2}{3}} = 2/7.$$

Thus the likelihood that the box offered was A is 2 /7.

Conditional probabilities arise in a natural way if we think of $\mathfrak{X}$ as a product space $\mathfrak{X}_1 x \mathfrak{X}_2$ in which a measure μ(not generally a product measure) is defined. Then if we write $(P(X_2/X_1)$ as the conditional probability that $x_2 \in X_2$ with respect to the condition $x_1 \in X_1$, we have

$$P(X_2/X_1) = \mu(X_1\ x\ X_2)/\mu_1(X_1)$$

where $$\mu_1(X_1) = \mu(X_1\ x\ \mathfrak{X}_2).$$

94

The set X_1, may reduce to a single point x_1, and the definition remains valid provided that $\mu_1(x_1) > 0$. But usually $\mu_1(x_1) = 0$, but the conditional probability with respect to a single point is not difficult to define. It follows from the Randon-Nikodym theorem that for fixed $X_2, \mu(X_1 x X_2)$ has a Radon derivative which we can write $R(x_1, X_2)$ with the property that

$$\mu(X_1\ x\ X_2) = \int_{x_1} R(x_1, X_2) d\mu_1$$

for all measurable X_1. For each $X_2, R(x_1, X_2)$ is defined for almost all x_1 and plainly $R(x_1, \mathfrak{X}_2) = 1$ p.p. But unfortunately,since the number of measurable sets X_2 is not generally countable, the union of all the exceptional sets may not be a set of measure zero. This means that we cannot assume that, for almost all $x_1, R(x_1, X_2)$ is a measure defined on all measurable sets X_2. If how ever, it is , we write it $P(X_1/x_1)$ and call it that conditional probability that $x_2 \mathcal{E} X_2$ subject to the condition that x_1 has a specified value.

Suppose now that (x_1, x_2) is a random variable in the plane with probability density $f(x_1, x_2)$($i.e\, f(x_1, x_2) \geq 0$ and $\iint f(x_1, x_2) dx_1 dx_2 = 1$). Then we can define *conditional probability densities* as follows:

$$P(x_2/x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} f_1(x_1) = \int\limits_{-\infty}^{\infty} f(x_1, x_2) dx_2$$

provided that $f_1(x_1) > 0$. **95**

The *conditional expectation* of x_2 for a fixed value of x_1 is

$$m(x_1) = \int\limits_{-\infty}^{\infty} x_2 P(x_2/x_1) dx_2 = \frac{\int\limits_{-\infty}^{\infty} x_2 f(x_1, x_2) dx_2}{\int\limits_{-\infty}^{\infty} f(x_1, x_2) dx_2}$$

The conditional standard deviation of x_2 for the value x_1 is $\sigma(x_1)$ where

$$\sigma^2(x_1) = \int\limits_{-\infty}^{\infty} (x_2 - m(x_1))^2 P(x_2/x_1) dx_2$$

$$\frac{\int\limits_{-\infty}^{\infty} (x_2 - m(x_1))^2 f(x_1, x_2) dx_2}{\int\limits_{-\infty}^{\infty} f(x_1, x_2) dx_2}$$

The curve $x_2 = m(x_1)$ is called the *regression curve of* x_2 *on* x_1. It has following minimal property it gives the least value of

$$E(x_2 - g(x_1))^2) = \iint\limits_{RXR} (x_2 - g(x_1))^2 f(x_1, x_1) dx_1 dx_2$$

for all possible curves $x_2 = g(x_1)$. If the curves $x_2 = g(x_1)$ are restricted to specified families, the function which minimizes E in that family. For example, the linear regression is the line $x_2 = Ax_1 + B$ for which **96**
$E((x_2 - Ax_1 - B)^2)$ is least, the n^{th} degree polynomial regression is the polynomial curve of degree n for which the corresponding E is least.

9. Sequences of Random Variables

We can define limit processes in connection with sequences of random variables in several different ways. The simplest is the convergence of the distribution or characteristic functions, $F_n(x)$ or $\phi_n(t)$, of random real numbers x_n to a limiting distribution or characteristic function $F(x)$ or $\phi(t)$. As in Theorem 15, it is sufficient to have $\lim_{n\longrightarrow\infty} F_n(x) = F(x)$ at every continuity point of $F(x)$. Note that this does not involve any idea of a limiting random variable. If we wish to introduce this idea, we must remember that it is necessary, when making probability statements about two or more random variables, to specify their joint probability distribution in their product space.

There are two important definitions based on this idea. We say that a sequence of random variables x_n *converges in probability* to a limit random variable x, and write

$$x_n \longrightarrow x \text{ in prob.}$$

if

$$\lim_{n\longrightarrow\infty} P(|x_n - x| > \epsilon) = 0$$

for every $\epsilon > 0$, P being the *joint* probability in the product space of x_n and x. In particular, if C is a constant, $x_n \longrightarrow C$ in prob, if

$$\lim_{n\rightarrow\infty} E(| x_n - x |^{\alpha}) = 0.$$

97

The most important case is that in which $\alpha = 2$. The following result is got almost immediately from the definition.

Theorem 22. *If $F_n(X)$ is the distribution function of x_n the necessary and sufficient condition that $x_n \rightarrow 0$ in prob. is that*

$$F_n(x) \rightarrow D(x)$$

where $D(x) = 0$, $x < 0; = 1, x \geq 0$ is the singular distribution. The necessary and sufficient condition that $x_n \to 0$ in mean of order α is that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} |\, x \,|^{\alpha} \, dF_n(x) = 0.$$

Theorem 23. (i) *If $x_n \to x$ in prob., then $F_n(x) \to F(x)$.*

(ii) *If $x_n \to x$ in mean, then $x_n \to x$ in prob. and $F_n(x) \to F(x)$. As corollaries we have*

Theorem 24 (Tchebycheff)**.** *If x_n has mean $\overline{x}_n$ and standard deviation σ_n then $x_n - \overline{x}_n \to 0$ in prob. if $\sigma_n \to 0$.*

Theorem 25 (Bernoulli: Weak law of large numbers)**.** *If $\xi_1, \xi_2, \ldots..$ are independent random variables with means $\overline{\xi}_1, \overline{\xi}_2, \ldots$ and standard deviations $\sigma_1, \sigma_2, ..$ and if*

$$x_n = \frac{1}{n}\sum_{\nu=1}^{n} \xi_\nu m_n = \frac{1}{n}\sum_{\nu=1}^{n} \overline{\xi}_\nu$$

then $$X_n - m_n \to 0 \text{ in prob. if } \sum_{\nu=1}^{n} \sigma_\nu^2 = 0(n^2)$$

98

Theorem 26 (Khintchine)**.** *If ξ_ν are independent random variables with the same distribution function and finite mean m, then*

$$x_n = \frac{1}{n}\sum_{\nu=1}^{n} \xi \to m \text{ in prob.}$$

(Note that this cannot be deduced from Theorem 25 since we do not assume that ξ_ν has finite standard deviation.)

Proof. Let $\phi(t)$ be the characteristic function of ξ_ν so that the characteristic function of x_n is $(\phi(t/n))^n$. If

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x),$$

we have $$\phi(t) - 1 - mit = \int_{-\infty}^{\infty} \left(e^{itx} - 1 - itx\right) dF(x).$$

Now $\left|\frac{e^{itx}-1-itx}{t}\right|$ is majorised by a multiple of $| x |$ and $\to 0$ as $t \to 0$ for each x.

Hence, by Lebesgue's theorem,

$$\phi(t) - 1 - \text{mit} = \sigma(t) \text{ as } t \to 0.$$

Thus

$$(\phi(t/n))^n = \left[1 + \frac{\text{mit}}{n} + 0\left(\frac{1}{n}\right)\right]^n \to e^{\text{mit}} asn \to \infty,$$

and since e^{mit} is the characteristic function of $D(x - m)$ the conclusion follows easily. □

99 In these definitions we need only the joint distribution of x and each x_N separately. In practice, of course, we may know the joint distributions of some of the x_n s (they may be independent, for example), but this is not necessary.

On the other hand, when we come to consider the notion of a random sequence, the appropriate probability space is the infinite product space of all the separate variables. This is a deeper concept than those we have used till now and we shall treat it later as a special case of the theory of random functions.

10. The Central Limit Problem

We suppose that

$$x_n = \sum_{\nu} x_{n\nu}$$

is a finite sum of independent random real variables x_{n_ν} and that $F_{n_\nu}(x)$, $F_{n\nu}(x)$, $\phi_{n_\nu}(t)$, $\phi_{n\nu}(t)$ are the associated distribution and characteristic functions. The general central limit problem is to find conditions under which $F_n(x)$ tends to some limiting function $F(x)$ when each of the components x_{n_ν} is small (in a sense to be defined later) in relation to x_n. Without the latter condition, there is no general result of this kind. Theorems 25 and 26 show that $F(x)$ may take the special form $D(x)$ and the next two theorems show that the Poisson and normal forms are also admissible. The general problem includes that of finding the most general class of such functions. The problem goes back to Bernoulli and Poisson and was solved (in the case of R) by Khintchine and P. Lévy.

Theorem 27 (Poisson)**.** *The binomial distribution* $P(x = \nu) = \binom{n}{\nu} p^\nu q^{n-\nu}$, **100**
$p = -\frac{c}{n}$, $q = 1 - p$, *c constant, tends to the Poisson distribution with mean c as* $n \to \infty$.

Proof.

$$\begin{aligned} \varphi_n(t) &= (q + pe^{it})^n \\ &= \frac{[1 + c(e^{it} - 1)]^n}{n} \\ &\to e^{c(e^{it}-1)} \end{aligned}$$

which, after Theorem 16, is sufficient. □

Theorem 28 (De Moivre)**.** *If* ξ_1 , $\xi_2 \cdots$ *are independent random variables with the same distribution, having mean 0 and finite standard deviation* σ*, then the distribution of*

$$x_n = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{\sqrt{n}}$$

tends to the normal distribution (0, σ^2*).*

This is proved easily using the method of Theorem 26.

The general theory is based on a formula due to Khintchine and Levy, which generalizes an earlier one for distributions of finite variance due to Kolmogoroff.

We say that $\psi(t)$ is a $K - L$ function with representation (a, G) if

$$\psi(t) = iat + \int_{-\infty}^{\infty} \left[e^{itx} - 1 - \frac{itx}{1 + x^2} \right] \frac{1 + x^2}{x^2} dG(x)$$

where a is a real number and $G(x)$ is bounded and non-decreasing in (−∞ , ∞). The value of the integrand at $x = 0$ is defined to be $-\frac{1}{2}t^2$ and it is then continuous and bounded in $-\infty < x < \infty$ for each fixed t. **101**

Theorem 29. *A $K - L$ function $\psi(t)$ is bounded in every finite interval and defines a, G uniquely.*

Proof. The first part is obvious.

If we define

$$\theta(t) = \psi(t) - \frac{1}{2}\int_{-1}^{1}\psi(t+u)du = \frac{1}{2}\int_{-1}^{1}(\psi(t) - \psi(t+u))du,$$

we have

$$\theta(t) = \frac{1}{2}\int_{-1}^{1}du\int_{-\infty}^{\infty}\left(e^{itx}(1 - e^{iux})\frac{iux}{1+x^2}\right)\frac{1+x^2}{x^2}dG(x)$$

$$= \int_{-\infty}^{\infty}e^{itx}\left(1 - \frac{\operatorname{Sin} x}{x}\right)\frac{1+x^2}{x^2}dG(x) = \int_{-\infty}^{\infty}e^{itx}dT(x)$$

where

$$T(x) = \int_{-\infty}^{x}\left[1 - \frac{\sin\ y}{y}\right]\frac{1+y^2}{y^2}dG(y),$$

$$G(x) = \int_{-\infty}^{x}\frac{y^2}{(1 - siny/y)(1+y^2)}dT(x)$$

since $\left[1 - \frac{\sin y}{y}\right]\frac{1+y^2}{y^2}$ and its reciprocal are bounded. □

This proves the theorem since $G(x)$ is defined uniquely by $T(x)$ which is defined uniquely $\psi(t)$ by Theorem 14 and this in turn is defined uniquely by $\psi(t)$.

102 The next theorem gives analogues of theorems 15 and 16. We shall write $G_n \to G$ if $G_n(x)$ and $G(x)$ are bounded and increasing and $G_n(x)$ $G(x)$ at every continuity point of $G(x)$ and at ±∞

Theorem 30. *If $\psi_n(t)$ has $K - L$ representation (a_n, G_n) for each n and if $a_n \to a$, $G_n \to G$ where $G(x)$ is non-decreasing and bounded, then $\psi_n(t) \to \psi(t)$ uniformly in any finite interval.*

Conversely, if $\psi_n(t) \to \psi(t)$ for all t and $\psi(t)$ is continuous at 0, then $\psi(t)$ has a $K - L$ representation (a, G) and $a_n \to a$, $G_n \to G$.

Proof. The first part is proved easily using the method of Theorem 15. For the second part, define $\theta_n(t)$, $T_n(t)$ as in the last theorem. Then, since $\theta_n(t) \to \theta(t) = \psi(t) - \frac{1}{2}\int_{-1}^{1}\psi(t+u)du$, which is continuous at 0, it follows from Theorem 16 that there is a non-decreasing bounded function $T(x)$ such that $T_n \to T$. Then $G_n \to G$ where $G(x)$ is defined as in Theorem 29, and is bounded, and $\psi(t)$ plainly has $K - L$ representation (a, G) where $a_n \to a$. □

Definition. *We say that the random variables $x_{n1}, x_{n2}, \ldots$ are uniformly asymptotically negligible (u.a.n.) if, for every $\in > 0$,*

$$\sup_{\nu} \int\limits_{|x| \geq \in} dF_n(x) \to 0 \text{ as } n \to \infty.$$

The condition that the variables are u.a.n implies that the variables are "centered" in the sense that their values are concentrated near 0. In the general case of u.a.n variables by considering the new variables $x_{n_\nu} - C_{n_\nu}$. Thus, we need only consider the u.a.n case, since theorems for this case can be extended to the u.a.n. case by trivial changes. We prove an elementary result about u.a.n. variables first. **103**

Theorem 31. *The conditions*

(i) x_{n_ν} *are u.a.n.*

(ii) $\sup\limits_{\nu} \int\limits_{-\infty}^{\infty} \frac{x^2}{1+x^2} dF_{n_\nu}(x + a_{n_\nu}) \to 0$ *for every set of numbers* $\left(a_{n_\nu}^{-\infty}\right)$ *for which* $\sup\limits_{\nu} \left|a_{n\nu}\right| \to 0$ *as* $n \to \infty$ *are equivalent and each implies that*

(iii) $\sup\limits_{\nu} | \varphi_{n\nu}(t) - 1 | \to 0$ *as* $n \to \infty$, *uniformly in every finite* $t-$*interval.*

Proof. The equivalence of (i) and (ii) follows from the inequalities

$$\int_{-\infty}^{\infty} \frac{x^2}{1+x^2} dF_{n\nu}(x + a_{n_\nu}) \leq (\epsilon + | a_{n_\nu} |^2) + \int_{|x|\geq\epsilon} dF_{n_\nu}(x)$$

$$\int_{|x|\geq\epsilon} dF_{n_\nu}(x) \leq \frac{1+\epsilon^2}{\epsilon^2} \int_{|x|\geq\epsilon} \frac{x^2}{1+x^2} dF_{n_\nu}(x).$$

For (iii) we use the inequality $| 1 - e^{itx} | \leq | xt | \; if \; | xt | \leq 1$ and deduce that

$$| \varphi_{n\nu}(t) - 1 | = | \int_{-\infty}^{\infty} (e^{itx} - 1) dF_{n_\nu}(x) | \leq \epsilon | t | + 2 \int_{|x|\geq\epsilon} dF_{n_\nu}(x).$$

□

104 **Theorem 32** (The Central Limit Theorem). *Suppose that $x_{n\nu}$ are independent u.a.n. variables and that $F_n \to F$. Then.*

(i) $\psi(t) = \log \varphi(t)$ *is a $K - L$ function*

(ii) *If $\psi(t)$ has representation (a, G) and the real numbers a_n satisfy*

$$\int_{-\infty}^{\infty} \frac{x}{1+x^2} dF_{n_\nu}(x + a_{n_\nu}) = 0,$$

and are bounded uniformly in ν then

$$G_n(x) = \int_{-\infty}^{x} \frac{y^2}{1+y^2} dF_{n_\nu}(y + a_{n_\nu}) \to G(x),$$

$$a_n = \sum_\nu a_{n_\nu} \to a.$$

Conversely, if the conditions (i) and (ii) hold, then $F_n \to F$.

Proof. It follows from the definition of the $a_{n\nu}$ that

$$e^{-ita_{n_\nu}}\phi_{n_\nu}(t) = \int\limits_{-\infty}^{\infty} e^{itx} dF_n(x + a_{n\nu}) = 1 + \gamma_{n_\nu}(t)$$

where $$\gamma_{\sqrt{n_\nu}}(t) = \int\limits_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1 + x^2}\right) dF_{\sqrt{n_\nu}}(x + a_{\sqrt{n_\nu}}),$$

and

$$\alpha_{\sqrt{n_\nu}}(t) = -\Re(\gamma_{\sqrt{n_\nu}}(t)) \geq 0 \tag{1}$$

It follow easily from the u.a.n. condition and the definition of $a_{n\nu}$ that $a_{\sqrt{n_\nu}} \to 0$ uniformly in ν as $n \to \infty$ and from Theorem 31 that **105**
$\gamma_{n_\nu}(t) \to 0$ uniformly in ν for $| t |\leq H$ where $H > 0$ is fixed. Hence

$$\log \varphi_{n_\nu}(t) = ita_{n_\nu} + \gamma_{n_\nu}(t) + 0(| \gamma_{n_\nu}(t) |^2),$$

the 0 being uniform in ν and , by addition,

$$\log \varphi_n(t) = \text{ it } a_n + \sum_\nu \gamma_{n_\nu}(t) + 0[\sum | \gamma_{n_\nu}(t) |^2], \tag{2}$$

uniformly in $| t |\leq H$.

Now let

$$\begin{aligned} A_{\sqrt{n_\nu}} &= \frac{1}{2h}\int_{-H}^{H} \propto_{n_\nu} (t)dt \\ &= \int\limits_{-\infty}^{\infty}\left[1 - \frac{\sin Hx}{Hx}\right] dF_{n_\nu}(x + a_{n_\nu}) \end{aligned}$$

Using the inequality

$$| e^{itx} - 1 - \frac{itx}{1 + x^2} |\leq C(H)\left[1 - \frac{\sin Hx}{Hx}\right]$$

for $| t |\leq H$, we have

$$| \gamma_{n_\nu}(t) |\leq CA_{n_\nu},$$

and therefore , taking real parts in (2) and using the fact that $\sup\limits_{\nu} |\gamma_{n_\nu}(t)| \to 0$ uniformly in $|t| \leq H$,

$$\sum_{\nu} \alpha_{n_\nu}(t) \leq -\log|\varphi_n(t)| + \circ\left(\sum_{\nu} A_{n_\nu}\right)$$

This again holds uniformly in $|t| \leq H$, and after integration we get

$$\sum_{\nu} \geq -\frac{1}{2H}\int_{-H}^{H} \log|\varphi_n(t)|\, dt + \circ\left(\sum A_{n_\nu}\right)$$

106 from which it follows that $\sum_\nu A_{n_\nu} = 0(1)$ and that

$$\log \varphi_n(t) = i\,t\,a_n + \sum_{\nu} \gamma_{n_\nu}(t) + 0(1), \tag{3}$$

uniformly for $|t| \leq H$, and the before, since H is at our disposal, for each real t.

The first part of the conclusion follows from Theorem 30.

For the converse, our hypothesis implies that $G_n(\infty) \to G(\infty)$ and if we use the inequality

$$\left|e^{itx} - 1 - \frac{itx}{1+x^2}\right| \leq C(t)\frac{x^2}{1+x^2},$$

it follows from (1) that

$$\sum_{\nu} |\gamma_{n_\nu}(t)| \leq C(t) G_n(\infty) = 0(1)$$

uniformly in γ. But $\gamma_\nu(t) \to 0$ uniformly in γ for any fixed t so that (2) and (3) remain valid and

$$\begin{aligned}\log \varphi_n(t) \to ita + \int_{-\infty}^{\infty} (e^{itx} - 1 - \frac{itx}{1+x^2})\frac{1+x^2}{x^2}\, dG(x)\\ = \psi(t) \log\ \varphi(t,)\end{aligned}$$

since $G_n \to G$. Hence $\varphi_n(t) \to \varphi_(t)$ as we require. □

Notes on Theorem 32.

(a) The first part of the theorem shows that the admissible limit functions for sums of u.a.n variables are those for which $\log \varphi(t)$ is a $K - L$ function.

(b) The numbers a_{n_ν} defined in stating the theorem always exist when 107
n is large since $\int_{-\infty}^{\infty} \frac{x}{1+x^2} dF_{n_\nu}(x+\xi)$ is continuous in ξ and takes positive and negative values at $\xi = \mp 1$ when n is large. They can be regarded as extra correcting terms required to complete the centralization of the variables. The u.a n. condition centralizes each of them separately, but this is not quite enough.

(c) The definition of a_{n_ν} is not the only possible one. It is easy to see that the proof goes through with trivial changes provided that the a_{n_ν} are defined so that $a_{n_\nu} \to 0$ and

$$\int_{-\infty}^{\infty} \frac{x}{1+x^2} dF_{n_\nu}(x + a_{n_\nu}) = 0\left(\int_{-\infty}^{\infty} \frac{x^2}{1+x^2} dF_{n_\nu}(x + a_{n_\nu})\right)$$

uniformly in ν as $n \to \infty$, and this is easy to verify if we define a_{n_ν} by

$$a_{n_\nu} = \int_{-\tau}^{\tau} x\, dF_n(x)$$

for some fixed $\tau > 0$. This is the definition used by Gnedenko and Lévy.

WE can deduce immediately the following special cases.

Theorem 33 (Law of large numbers)**.** *In order that F_n should tend to the singular distribution with $F(x) = D(x - a)$ it is necessary and sufficient that*

$$\sum_\nu \int_{-\infty}^{\infty} \frac{y^2}{1+y^2}\, d\, F_{n_\nu}(y + a_{n_\nu}) \to 0, \sum_\nu a_{n_\nu} \to a$$

(Here $\psi(t) = ita$, $G(x) = 0$).

108 **Theorem 34.** *In order that F_n should tend to the Poisson distribution with parameter c, it is necessary and sufficient that*

$$\sum_{\nu} a_{n_\nu} \to \frac{1}{2}\, c \text{ and } \sum_{\nu} \int_{-\infty}^{x} \frac{y^2}{1+y^2}\, d\, F_{n_\nu}(y + a_n) \to \frac{c}{2}\, D(x-1)$$

(Here $\psi(t) = c(e^{it} - 1)$, $a = \frac{1}{2}c$, $G(x) = \frac{c}{2}\, D(x-1)$)

Theorem 35. *In order that F_n should tend to the normal (α, σ^2) distribution, it is necessary and sufficient that*

$$\sum_{\nu} a_{n_\nu} \to \alpha \text{ and } \int_{-\infty}^{x} \frac{y^2}{1+y^2} dF_{n\nu}(y + a_{n_\nu}) \to \sigma^2 D(x).$$

(Here $\psi(t) = i\, t\alpha - \frac{\sigma^2 t^2}{2}$, $a = \alpha$, $G(x) = \sigma^2 D(x)$. From this and the note (c) after Theorem 33, it is easy to deduce

Theorem 36 (Liapounoff)**.** *If $x_{n\nu}$ has mean o and finite variance $\sigma^2_{n\nu}$ with $\sum_\nu \sigma^2_{n\nu} = 1$ a necessary and sufficient condition that x_n should tend to normal $(0, 1)$ distribution is that for every $\in > 0$,*

$$\sum_{\nu} \int_{|x| \geq \in} x^2\, d\, F_{n\nu}\,(x) \to 0 \text{ as } n \to \infty.$$

The distributions for which log $\varphi(t)$ is a K – L function can be characterized by another property. We say that a distribution is infinitely divisible (i.d.) if, for every n we can write

$$\phi(t) = (\phi_n(t))^n$$

109 *where $\phi_n(t)$ is a characteristic function. This means that it is the distribution of the sum of n independent random variables with the same distribution.*

Theorem 37. *A distribution is i.d. if and only if $\log \phi(t)$ is a $K - L$ function.*

This follows at once from Theorem 32. That the condition that a distribution be i.d. is equivalent to a lighter one is shown by the following

Corollary 1. *$\phi(t)$ is i.d. if there is a sequence of decompositions (not necessarily with identically distributed components) in which the terms are u.a.n.*

Corollary 2. *If a distribution is i.d., $\phi(t) \neq 0$.*

Theorem 38. *A distribution is i.d. if and only if it is the limit of finite compositions of u.a.n. Poisson distributions.*

Proof. The result is trivial in one direction. In the other, we observe that the integral

$$\int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2} dG(x)$$

can be interpreted as a Riemann - Stieltjes integral and is the limit of finite sums of the type

$$\sum_j b_j \left(e^{it\xi_j} - 1 - \frac{it\xi_j}{1+\xi_j^2}\right)$$

each term of which corresponds to a Poisson distribution. □ **110**

11. Cumulative sums

Cumulative sums

$$x_n = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}$$

in which $\xi_1, \xi_2, \ldots \xi_n$ are independent and have distribution functions $B_1(x), B_2(x), \ldots, B_n(x)$ and characteristic functions $\beta, (t), \beta(t), \ldots, \beta_n(t)$ are included in the more general sums considered in the central limit theorem. It follows that the limiting distribution of x_n is i.d. and if $\varphi(t)$ is the limit of the characteristic functions $\phi_n(t)$ of x_n, then $\log \varphi(t)$ is a $K-L$ function function. These limits form, however, only a proper subclass of the $K - L$ class and the problem of characterizing this subclass was proposed by Khint chine and solved by Lévy. We denote the class by L.

As in the general case, it is natural to assume always that the components $\frac{\xi_\nu}{\lambda_n}$ are u.a.n.

Theorem 39. *If ξ_ν/λ_n are u. a. n. and $\phi_n(t) \to \phi(t)$ where $\phi(t)$ is non-singular, then $\lambda_n \to \infty$, $\frac{\lambda_{n+1}}{\lambda_n} \to 1$.*

Proof. Since $\lambda_n > 0$, either $\lambda_n \to \infty$ or (λ_n) contains a convergent subsequence (λ_{n_k}) with limit λ. The u.a.n. condition implies that $\beta_\nu\left(\frac{t\lambda}{\lambda_n}\right) \to 1$ for every t and therefore, by the continuity of $\beta_\nu(t)$,

$$\beta_\nu(t) = \lim_{k\to\infty} \beta_\nu\left(\frac{t\lambda}{\lambda_{n_k}}\right) = 1$$

111 for all t. This means that every $\beta_\nu(t)$ is singular, and this is impossible as $\varphi(t)$ is not singular.

For the second part, since

$$\frac{\lambda_n x_n}{\lambda n + 1} = \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{\lambda_{n+1}} = x_{n+1} - \frac{\xi_{n+1}}{\lambda_{n+1}}$$

and the last term is asymptotically negligible, $\frac{\lambda_n x_n}{\lambda_{n+1}}$ and x_{n+1} have the same limiting distribution $F(x)$, and therefore

$$F_n\left(\frac{x\lambda_{n+1}}{\lambda_n}\right) \to F(x), F_n(x) \to F(x)$$

Now if $\frac{\lambda_{n+1}}{\lambda_n} = \theta_n$, we can choose a subsequence (θ_{n_k}) which either tends to ∞ or to some limit $\theta \geq 0$. In the first case $F_{n_k}(x\theta_{n_k}) \to F(\pm\infty) \neq F(x)$ for some x. In the other case

$$F(x) = \lim_{k\to\infty} F_{n_k}(\theta_{n_k} x) = F(\theta x)$$

whenever x and θx are continuity points of $F(x)$ and this is impossible unless $\theta = 1$.

A characteristic function $\phi(t)$ is called *self-decomposable* (s.d) if, for every c in $0 < c < 1$ it is possible to write

$$\phi(t) = \phi(ct)\phi_c(t)$$

112 where $\varphi_c(t)$ is a characteristic function □

Theorem 40 (P.Lévy)**.** *A function $\varphi(t)$ belongs to L if and only if it is self-decomposable , and $\varphi_c(t)$ is then i.d*

Proof. First suppose that $\varphi(t)$ is s.d. if it has a positive real zero, it has a smallest, $2a$, since it is continuous, and so

$$\varphi(2a) = 0, \varphi(t) \neq 0 for 0 \leq t < 2a.$$

Then $\varphi(2ac) \neq 0$ if $0 < c < 1$, and since $\varphi(2a) = \varphi(2ac)\varphi_c(2a)$ it follows that $\varphi_c(2a) = 0$. Hence

$$\begin{aligned} 1 = 1 - \Re(\varphi_c(2a)) &= \int_{-\infty}^{\infty} (1 - \cos 2ax) dF_c(x) \\ &= 2 \int_{-\infty}^{\infty} (1 - \cos ax)(1 + \cos ax) dF_c(x) \leq 4 \int_{-\infty}^{\infty} (1 - \cos ax) dF_c(x) \\ &= 4(1 - \Re(\varphi_c(a))) = 4(1 - \Re(\varphi(a) 1 \varphi(ca))) \end{aligned}$$

This leads to a contradiction since $\varphi(ca) \to \phi(a)$ as $c \to 1$, and it follows therefore that $\varphi(t) \neq 0$ for $t \geq 0$ and likewise for $t < 0$.

If $1 \leq \mathcal{V} \leq n$ it follows from our hypothesis that

$$\beta_{\mathcal{V}}(t) = \varphi_{\frac{v-1}{v}}(\mathcal{V}t) \frac{\varphi(\mathcal{V}t)}{\varphi((\mathcal{V}-1)t)}$$

is a characteristic function and the decomposition

$$\varphi(t) = \varphi_n(t) = \prod_{=1}^{n} \beta \mathcal{V}(t/n)$$

shows that $\varphi(t)$ is of type L with $\lambda_n = n$

Conversely if we suppose that $\varphi(t)$ is of type L we have

$$\begin{aligned} \varphi_n(t) &= \prod_{r=1}^{n} \beta_r(t\lambda_n) \\ \varphi_{n+m}(t) &= \prod_{v=1}^{n+m} \beta_\gamma(t/\lambda_{n+m}) = \varphi_n(\lambda_n t/\lambda_{n+m}) \chi_{n,m}(t), \end{aligned}$$

where

$$\chi_{n,m}(t) = \prod_{v=n+1}^{n+m} \beta_\gamma(t/\lambda_{n+m})$$

Using theorem 39, we can those n, $m(n) \to \infty$ so that $\lambda_n/\lambda_{n+m} \to c(0 < c < 1)$ and then $\varphi_{n+m}(t) \to \varphi(t)$. Since $\varphi_n(t) \to \varphi(t)$ uniformly in any finite t-interval $\varphi_n(\lambda_n t/\lambda_{n+m}) \to \varphi(ct)$. It follows that $\chi_{n,m}(t)$ has a limit $\phi_c(t)$ which is continuous at $t = 0$ and is therefore a characteristic function by theorem 16. Moreover the form of $\chi_{n,m}(t)$ shows that $\varphi_c(t)$ is i.d.

The theorem characterizes L by a property of $\varphi(t)$. It is also possible to characterize it by a property of $F(x)$. □

Theorem 41 (P.Lévy)**.** *The function $\varphi(t)$ of L are those for which* $\log \varphi(t)$ *has a $K-L$ representation (a, G) in which $\frac{x^2+1}{x}G'(x)$ exists and decreases outside a countable set of points.*

Proof. If we suppose that $\varphi(t)$ is of class L and $0 < c < 1$ and ignore terms of the form iat we have

$$\log \varphi_c(t) = \int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2} dG(x)$$
$$- \int_{-\infty}^{\infty} \left[e^{itx} - 1 - \frac{itc^2 x}{c^2 + x^2}\right] \frac{x^2 + c^2}{x^2} dG(x/c)$$

114 and the fact that $\varphi_c(t)$ is i.d. by Theorem 40 implies that $Q(x) - Q(bx)$ decreases, where $b = 1/c > 1$, $x > 0$ and

$$Q(x) = \int_x^{\infty} \frac{y^2 + 1}{y^2} dG(y)$$

If we write $q(x) = Q(e^x)$ this means that $q(k) - q(x + d)$ decrease for $x > 0$ if $d > 0$. If follows that $q'(x)$ exists and decreases outside a countable set (see for instance Hardy, Littlewood, Polya: Inequalities , P.91 together with Theorem 48 Chapter I on page 51)

Then since

$$\frac{x^2}{x^2+1}\frac{Q(x)-Q(x+h)}{h} \leq \frac{G(x+h)-G(x)}{h}$$
$$\leq \frac{(x+h)^2}{(x+h)^2+1}\frac{Q(x)-Q()x+h}{h}$$

We have $G'(x) = \frac{x^2}{x^2+1}Q'(x)$ and $\frac{x^2+1}{x}G'(x) = xQ'(x)$ which also exists and decreases outside a countable set. The same argument applies for $x < 0$. The converse part is trivial.

A more special case arises if we suppose that the components are identically distributed and the class L^* of limits for sequences of such cumulative sums can again be characterized by properties of the limits $\varphi(t)$ or $G(x)$

We say that $\varphi(t)$ is *stable*, if for every positive constant b, we can find constants a, b' so that

$$\varphi(t)\varphi(bt) = e^{iat}\varphi(b^{'}t)$$

This implies, of course, that φ (t) is s.d. and that φ_c has the form $e^{ia't}\varphi(c't)$. □

Theorem 42 (P.Lévy)**.** *A characteristic function $\varphi(t)$ belongs to L^* if and only if it is stable.* **115**

Proof. If $\varphi(t)$ is stable, we have on leaving out the inessential factors of the form $e^{i\alpha t}$, $(\varphi(t))^n = \varphi(\lambda_n t)$ for some $\lambda_n > 0$ and so

$$\varphi(t) = (\varphi(t/\lambda_n))^n = \prod_{\nu=1}^{n}\beta_\nu(t/\lambda_n) \text{ with } \beta_\nu(t) = \varphi(t),$$

which is enough to show that $\varphi(t)$ belongs to L^*.

Conversely, if we suppose that a sequence λ_n can be found so that

$$\varphi_n(t) = (\varphi(t/\lambda_n))^n \to \varphi(t),$$

we write $n = n_1 + n_2$,

$$\varphi_n(t) = (\varphi(t/\lambda_n))^{n_1}(\varphi(t/\lambda_n))^{n_2} = \varphi_{n_1}(t\lambda_{n_1}/\lambda_n)\varphi_{n_2}(t\lambda_{n_2}/\lambda_n).$$

Then, if $0 < c < 1$, we choose n_1 so that $\lambda_{n_1}/\lambda_n \to c$ and it follows that $\varphi_{n_1}(t\lambda_{n_1}/\lambda_n) \to \phi(ct)$ and $\varphi_{n_2}(t\lambda_{n_2}/\lambda_n) \to \varphi_c(t)$. It is easy to show that this implies that $\varphi_c(t)$ has the form $e^{ia''t}\phi(c't)$.

It is possible to characterize the stable distributions in terms of log $\varphi(t)$ and $G(x)$. □

Theorem 43 (P.Lévy). *The characteristic function $\varphi(t)$ is stable if and only if*

$$\log \phi t = iat - A \mid t \mid^{\alpha} (1 + \frac{i\theta t}{t} \tan \frac{\pi\alpha}{2})$$

$$0 < \alpha < 1 \text{ or } 1 < \alpha < 2$$

or $\log \varphi(t) = iat - A \mid t \mid (1 + \frac{i\theta t}{t} \frac{2}{\pi} \log \mid t \mid)$ with $A > 0$ and $-1 \leq \theta \leq +1$.

116

Corollary. *The real stable distributions are given by $\varphi(t) = e^{-A\,|t|^{\alpha}}(0 < \alpha \leq 2)$.*

Proof. In the notation of Theorem 41, the stability condition implies that, for every $d > 0$, we can define d' so that

$$q(x) = q(x + d) + q(x + d')(x > 0).$$

Since $q(x)$ is real, the only solutions of this difference equation, apart from the special case $G(x) = AD(x)$ are given by

$$q(x) = A_1 e^{-\alpha x}, Q(x) = A_1 x^{-\alpha}, G'(x) = \frac{A_1 x^{1-\alpha}}{1 + x^2}, x > 0$$

and α satisfies $1 = \mathrm{e}^{-\alpha \mathrm{d}} + e^{-\alpha d'}$. We can use a similar argument for x <0 and we have also

$$q(x) = A_2 e^{-\alpha|\mathrm{x}|}, Q(x) = A_2 \mid \mathrm{x} \mid^{-\alpha}, G'(x) = \frac{A_2 \mid \mathrm{x} \mid^{1-\alpha}}{1 + x^2}, x < 0$$

where d, d', α are the same. Moreover, since $G(x)$ is bounded, we must have $0 < \alpha \leq 2$, and since the case $\alpha = 2$ arises when $G(x) = AD(x)$ and the distribution is normal, we can suppose that $0 < \alpha < 2$. Hence

$$\log\varphi(t) = iat + A_1\int_0^\infty \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{dx}{x^{\alpha+1}} + A_2\int_{-\infty}^0 \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{dx}{|\,x\,|^{\alpha+1}}$$

The conclusion follows, if $\alpha \neq 1$ form the formula **117**

$$\int_0^\infty (e^{itx} - 1)\frac{dx}{x^{\alpha+1}} = |\,t\,|^\alpha\ e^{-\alpha\pi i/2}\Gamma(-\alpha) \text{ if } 0 < \alpha < 1,$$

$$\int_0^\infty \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{dx}{x^{\alpha+1}} = |\,t\,|^\alpha\ e^{-\alpha\pi i/2}\Gamma(-\alpha) \text{ if } 1 < \alpha < 2$$

(easily proved by contour integration), Since the remaining components $\frac{itx}{1+x^2}$ or $\frac{itx}{1+x^2} - tx$ merely add to the term iat. If $\alpha = 1$, we use the formula

$$\int_0^\infty \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{dx}{x^2} = -\frac{\pi}{2}\,|\,t\,| - \text{it}\ \log|\,t\,| + ia_1 t,$$

which is easy to Verify. □

12. Random Functions

In our discussion of random functions, we shall not give proofs for all theorems, but shall content ourselves with giving references, in many cases.

Let Ω be the space of functions $x(t)$ defined on some space T and taking values in a space $\mathfrak{X}$. Then we call $x(t)$ a *random function* (or *process*) if a probability measure is defined in Ω. We shall suppose here that $\mathfrak{X}$ is the space of real numbers and that T is the same space or some subspace of it.

The basic problem is to prove the existence of measures in Ω with certain properties - usually that certain assigned sets in Ω are measurable and have assigned measures. These sets are usually associated with **118**
some natural property of the functions $x(t)$. It is sometimes convenient

to denote the function (which is a point Ω) by ω and the value of the function at t by $x(t, \omega)$..

A basic theorem is

Theorem 44 (Kolomogoroff)**.** *Suppose that for every finite set of distinct real numbers $t_1, t_2, \ldots, t_n$ we have a joint distribution function-$F_{t_1,t_2,\ldots,t_n}(\xi_1, \xi_2, \ldots, \xi_n)$ in R_n and that these distribution functions are consistent in the sense that their values are unchanged by like permutations of (t_i) and (ξ_i) and, if $n > m$,*

$$F_{t_1,t_2,\ldots,t_n}(\xi_1, \xi_2, \ldots, \xi_m, \infty, \ldots, \infty) = F_{t_1,t_2,\ldots,t_m}(\xi_1, \ldots, \xi_m).$$

Then a probability measure can be defined in Ω in such a way that

$$Prob(x(t_i) \leq \xi_i, i = 1, 2, \ldots, n) = F_{t_1,t_2,\ldots,t_n}(\xi_1, \ldots, \xi_n). \tag{1}$$

Proof. The set of functions defined by a finite number of conditions

$$a_i \leq x(t_i) \leq b_i$$

is called a *rectangular set* and the union of a finite number of rectangular sets is called a figure. It is plain that intersections of figures are also figures and that the system S_o of figures 1 and their complements is additive. Moreover, the probability measure μ defined in S_o by 1 is additive in S_o , and it is therefore enough, after Theorem 7 of Chapter
119 1, to show that μ is completely additive in S_o. It is enough to show that if I_n are figures and $I_n \downarrow 0$, then $\mu(I_n) \to 0$.

We assume that $\lim \mu(I_n) > 0$, and derive a contradiction. Since only a finite number of points t_i are associated with each I_n, the set of all these $t_i' s$ is countable and we can arrange them in a sequence (t_i). Now each I_n is the union of a finite number of the rectangular sets in the product space of finite number of the space of the variables $x_i = x(t_i)$ and we can select one of these rectangles, for $n = 1, 2, \ldots$ so that it contains a *closed* rectangle J_n with the property that $\lim\limits_{m \to \infty} \mu(J_n I_m) > 0$. Also we may choose the J_n so that $J_{n+1} \subset J_n$. We then obtain a decreasing sequence of closed non empty rectangles J_n defined by

$$a_{in} \leq y_i \leq b_{in} (i = 1, 2, \ldots, i_n)$$

For each i there is at least one point y_i which is contained in all the intervals $[a_{in}, b_{in}]$, and any function $x(t)$ for which $x(t_i) = y_i$ belongs to all I_n. This impossible since $I_n \downarrow 0$, and therefore we have $\mu(I_n) \to 0$.

As an important special case we have the following theorem on random sequences. ◻

Theorem 45. *Suppose that for every N we have joint distribution functions $F_N(\xi_1, \ldots, \xi_N)$ in R_N which are consistent in the sense of Theorem 44. Then a probability measure can be defined in the space of real sequences $(x_1, x_2, \ldots)$ in such a way that*

$$P(x_i \leq \xi_i i = 1, 2, \ldots, N)$$
$$= F_N(\xi_1, \ldots, \xi_N)$$

Corollary. *If $\{F_n(x)\}$ is a sequence of distribution functions, a probabil-* 120
ity measure can be defined in the space of real sequence so that if I_n are any open or closed intervals,

$$P(x_n \in I_n, n = 1, 2, \ldots, N) = \prod_{n=1}^{N} F_n(I_n).$$

The terms of the sequence are then said to be *independent*, and the measure is the product measure of the measures in the component spaces. The measures defined by Theorem 44 will be called K-measure. The probability measures which are useful in practice are generally extensions of K-measure, since the latter generally fails to define measures on important classes of functions. For example, if I is an interval, the set of functions for which $a \leq x(t) \leq b$ for t in I is not K-measurable.

In the following discussion of measures with special properties, we shall suppose that the basic K-measure can be extended so as to make measurable all the sets of functions which are used.

A random function $x(t)$ is called *stationary* (in the strict sense) if the transformation $x(t) \to x(t + a)$ preserves measures for any real a (or integer a in the case of sequences).

A random function $x(t)$ is said to have *independent increments* if the variables $x(t_i) - x(s_i)$ are independent for non-overlapping intervals (s_i, t_i). It is called *Gaussian* if the joint probability distribution 121 for $x(t_1), x(t_2), \ldots, x(t_n)$ for any finite set $t_1, t_2, \ldots, t_n$ is Gaussian in R_n. That is , if the functions F of Theorem 44 are all Gaussian. A random function $x(t)$ is called an L_2- function if it has finite variance for every t. This means that $x(t, w)$ belongs to $L_2(\Omega)$ as a function of w for each t, and the whole function is described as a *trajectory* in the Hilbert space $L_2(\Omega)$.

Many of the basic properties of an L_2-function can be described in terms of the *auto-correlation* function

$$\begin{aligned} r(s,t) &= E((x(s) - m(s))\overline{(x(t) - m(t))}) \\ &= E(x(s)\overline{x(t)}) - m(s)\overline{m(t)} \end{aligned}$$

where $m(t) = E((x(t))$.

A condition which is weaker than that of independent increments is that the increments should be *uncorrelated*. This is the case if $E(x(t) - x(s))\overline{(x(t') - x(s'))}) = E(x(t) - x(s))\overline{E(x(t') - x(s'))}$ for non- overlapping intervals (s, t), (s', t'). If an L_2-function is *centred* so that $m(t) = 0$ (which can always be done trivially by considering $x(t) - m(t)$), a function with uncorrelated increments has *orthogonal* increments, that is

$$E((x(t) - x(s))\overline{(x(t') - x(s'))}) = 0$$

for non-overlapping intervals. The function will then be called an *orthogonal random function*.

122 The idea of a *stationary* process can also be weakened in the same way. An L_2- function is *stationary in the weak sense* or stationary, if $r(s, t)$ depends only on $t - s$. We then write $\rho(h) = r(s, s + h)$.

We now go on to consider some special properties of random functions.

13. Random Sequences and Convergence Properties

The problems connected with random sequences are generally much simpler than those relating to random functions defined over a non-countable set. We may also use the notation w for a sequence and $x_n(w)$ for its n^{th} term.

Theorem 46 (The 0 or 1 principle: Borel, Kolmogoroff)**.** *The probability that a random sequence of independent variables have a property (e.g. convergence) which is not affected by changes in the values of any finite number of its terms is equal to 0 to 1.*

Proof. Let E be the set of sequences having the given property, so that our hypothesis is that, for every $N \geq 1$,

$$E = \mathfrak{X}_1 \ x \ \mathfrak{X}_2 \ x \ \ldots \ x \ \mathfrak{X}_n \ x \ E_N$$

where E_N is a set in the product space $\mathfrak{X}_{N+1} \ x \ \mathfrak{X}_{N+2} \ x \ \ldots$

It follows that if F is any figure, $F \ E = F \ x \ E_N$ for large enough N and

$$\mu(FE) = \mu(F)\mu(E_N) = \mu(F)\mu(E)$$

and since this holds for all figures F , it extends to measurable sets F. In **123**
particular, putting F=E, we get

$$\mu(E) = (\mu(E))^2, \mu(E) = 0or1.$$

We can now consider questions of convergence of series $\sum_{\nu=1}^{\infty} x_\nu$ of *independent* random variables. □

Theorem 47. *If* $s_n = \sum_{\nu=1}^{\infty} x_\nu \to s$ *p. p., then* $s_n - s \to 0$ *in probability and the distribution function of* s_n *tends to that of* s *. (This follows from Egoroff's theorem)*

Theorem 48 (Kolmogoroff's inequality)**.** *If* x_ν *are independent, with means 0 and standard deviations* σ_ν *and if*

$$T_N = \sup_{n \leq N} \mid s_n \mid, s_n = \sum_{\nu=1}^{n} x_\nu, \in > 0$$

then

$$P(T_N \geq \in) \leq \in^{\frac{1}{2}} \sum_{\nu=1}^{N} \sigma_\nu^2$$

Proof. Let

$$E = \in [T_N \geq \in] = \sum_{k=1}^{N} E_k$$

where $$E_k = \in [| s_k | \geq \in, T_{k-1} < \in]$$

It is plain that the E_k are disjoint . Moreover $\sum\limits_{\nu=1}^{N} \sigma_\nu^2 = \int\limits_{\Omega} s_N^2 d\mu$, since the x_ν are independent,

$$\begin{aligned}
&\geq \int_E s_N^2 d\mu = \sum_{k=1}^{N} \int\limits_{E_K} s_N^2 d\mu \\
&= \sum_{k=1}^{N} \int\limits_{E_K} (s_k + x_{k+1} + ... + x_N)^2 d\mu \\
&= \sum_{k=1}^{N} \int\limits_{E_K} s_k^2 d\mu + \sum_{k=1}^{N} \mu(E_k) \sum_{i=k+1}^{N} \sigma_i^2
\end{aligned}$$

124 since E_k involves only $x_1, \ldots, x_k$.

Therefore

$$\sum_{\nu=1}^{N} \sigma_\nu^2 \geq \sum_{k=1}^{N} \int\limits_{E_k} s_k^2 d\mu \geq \in^2 \sum_{k=1}^{N} \mu(E_k) = \in^2 \mu(E)$$

as we require. □

Theorem 49. *If x_ν are independent with means m_ν and $\sum\limits_{\nu=1}^{\infty} \sigma_\nu^2 < \infty$ then $\sum_1^\infty (x_\nu - m_\nu)$ converges p.p.*

Proof. It is obviously enough to prove the theorem in the case $m_\nu = 0$.

By theorem 48, if $\in > 0$

$$P\left(\sup_{|\leq n \leq N} | s_{m+n} - s_m | < \in\right) \geq \frac{1}{\in^2} \sum_{\nu=m+1}^{m+n} \sigma_\nu^2$$

and therefore

$$P\left(\sup_{n\geq 1} \mid s_{m+n} - s_m \mid > \in\right) \leq \frac{1}{\epsilon^2} \sum_{\nu=m+1}^{\infty} \sigma_\nu^2$$

and this is enough to show that

$$\lim_{m\to\infty} \sup_{n\geq 1} \mid s_{m+n} - s_m \mid = 0 p.p.$$

and by the general principle of convergence, s_n converges p.p. **125**

As a partial converse of this, we have □

Theorem 50. *If x_ν are independent with means m_ν and standard deviations $\sigma_\nu, \mid x_\nu \mid \leq c$ and $\sum_{=1}^{\infty} x_\nu$ converges in a set of positive measure (and therefore p.p. by Theorem 46), then $\sum_{\nu=1}^{\infty} \sigma_\nu^2$ and $\sum_{\nu=1}^{\infty} m_\nu$ converge.*

Proof. Let $\varphi_\nu(t), \vartheta(t)$ be the characteristic functions of x_ν and $s = \sum_{\nu=1}^{\infty} x_\nu$. Then it follows from Theorem 47 that

$$\prod_{\nu=1}^{\infty} \varphi_\nu(t) = \vartheta(t)$$

where $\vartheta(t) \neq 0$ in some neighbourhood of t=0, the product being uniformly convergent over every finite t-interval. Since

$$\varphi_\nu(t) = \int_{-c}^{c} e^{\text{itx}} dF_\nu(x)$$

it is easy to show that

$$\sigma_\nu^2 \leq -K \log \mid \varphi_\nu(t) \mid$$

if t is in some sufficiently small interval independent of ν, and it follows that $\sum_{\nu=1}^{\infty} \sigma_\nu^2 < \infty$. Hence $\sum_{\nu=1}^{\infty} (x_\nu - m_\nu)$ converges p.p. by Theorem 49, and since $\sum x_\nu$ converges p.p., $\sum m_\nu$ also converges. □

Theorem 51 (Kolmogoroff's three series theorem)**.** *Let x_ν be independent and $c > o$,*

$$x'_\nu = x_\nu if \mid x_\nu \mid \leq c$$
$$= 0 if \mid x \mid > c.$$

126 *Then $\sum\limits_1^\infty x_\nu$ converges p.p. if and only if the three series*

$$\sum_1^\infty P(\mid x_\nu \mid > c), \sum_1^\infty m_\nu^1, \sum_1^\infty \sigma_\nu'^2$$

converge, where m'_ν , σ'_ν are the means and standard deviations of the x'_ν.

Proof. First,if $\sum x_\nu$ converges p.p., $x_\nu \to 0$ and $x'_\nu = x_\nu$, $\mid x_\nu \mid < c$ for large enough ν for almost all sequences.

Let $p_\nu = P(\mid x_\nu \mid > c)$.

Now

$$\varepsilon\left[\lim_{\nu\to} \sup_{\infty} \mid x_\nu \mid < c\right] = \lim_{N\to\infty} \varepsilon[\mid x_\nu \mid < c] \text{ for } n \geq N]$$
$$= \lim_{N\to\infty} \bigcap_{\nu=N}^{\infty} \varepsilon[\mid x_\nu \mid < c].$$

Therefore

$$1 = P(\lim\sup \mid x_\nu \mid < c) = \lim_{N\to\infty} \prod_{\nu=N}^{\infty} (1 - p_\nu)$$

by the independence of the x_ν. Hence $\prod\limits_{\nu=1}^{\infty}(1 - p_\nu)$ so $\sum\limits_{\nu=1}^{\infty} p_\nu$ converge. The convergence of the other two series follows from Theorem 50.

127 Conversely, suppose that the three series converge, so that, by Theorem 50, $\sum\limits_1^\infty x'_\nu$ converges p.p. But it follows from the convergence of $\sum\limits_1^\infty p_\nu$ that $x_\nu = \mathfrak{X}'_\nu$ for sufficiently large ν and almost all series, and therefore $\sum\limits_1^\infty x_\nu$ also converges p.p. □

Theorem 52. *If x are independent, $s_n = \sum_{\nu=1}^{n} x_\nu$ converges if and only if $\prod_{\nu=1}^{\infty} \varphi_\nu(t)$ converges to a characteristic function.*

We do not give the proof. For the proof see j.L.Doob, Stochastic processes, pages 115, 116. If would seem natural to ask whether there is a direct proof of Theorem 52 involving some relationship between T_N in Theorem 48 and the functions $\varphi_\nu(t)$. This might simplify the whole theory.

Stated differently, Theorem 52 reads as follows:

Theorem 53. *If x_ν are independent and the distribution functions of s_n converges to a distribution function, then s_n converges p.p.*

This is a converse of Theorem 47.

Theorem 54 (The strong law of large numbers)**.** *If x_ν are independent, with zero means and standard deviations σ_ν, and if*

$$\sum_{\nu=1}^{n} \frac{\sigma_\nu^2}{\nu^2} < \infty$$

then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} x_\nu = 0 \ p.p.$$

Proof. Let $y\nu = \dfrac{x\nu}{\nu}$, so that y_ν has standard deviation σ_γ/ν. It follows **128**
then from Theorem 49 that $\sum y_\nu = \sum (x_\nu/\nu)$ converges p.p.

If we write $x_\nu = \sum_{j=1}^{\nu} x_j/j$,

$$x_\nu = \nu X_\nu - \nu X_{\nu=1},$$

$$\frac{1}{2}\sum_{\nu=1}^{n} x_\nu = \frac{1}{n}\sum_{\nu=1}^{n} (\nu X_\nu - \nu X_{\nu=1})$$

$$= X_n - \frac{1}{n}\sum_{\nu-1}^{n} X_{\nu-1}$$

$$= 0(1)$$

if X_n converges, by the consistency of (C, 1) summation.

If the series $\sum\limits_{1}^{\infty} x_\nu$ does not converge p.p. it is possible to get results about the order of magnitude of the partial sums $s_n = \sum\limits_{1}^{n} x_\nu$. The basic result is the famous *law of the iterated logarithm* of Khintchine. □

Theorem 55 (Khintchine; Law of the iterated logarithm)**.** *Let x_ν be independent, with zero means and standard deviations σ_ν*

Let

$$B_n = \sum_{\nu=1}^{n} \sigma_\nu^2 \longrightarrow \infty \text{ as } n \longrightarrow \infty.$$

129 *Then*

$$\limsup_{n\to\infty} \frac{| s_n |}{\sqrt{(2B_n \log\log B_n)}} = 1 p.p.$$

Corollary. *If x_ν have moreover the same distribution, with $\sigma_\nu = \sigma$ then*

$$\limsup_{n\to\infty} \frac{| s_n |}{\sqrt{(2n \log\log n)}} = \sigma \; p.p.$$

We do not prove this here. For the proof, see M. Loeve: Probability Theory, Page 260 or A. Khintchine : Asymptotische Gesetzeder Wahrsheinlichkeit srechnung, Page 59.

14. Markoff Processes

A random sequence defines a *discrete Markoff process* if the behaviour of the sequence x_ν for $\nu \geq n$ depends only on x_n (see page 123). It is called a *Markoff chain* if the number of possible values (or states) of x_ν is finite or countable. The states can be described by the *transition probabilities* ${}_np_{ij}$ defined as the probability that a sequence for which $x_n = i$ will have $x_{n+1} = j$. Obviously

$${}_np_{ij} \geq 0, \sum_J {}_np_{ij} = 1.$$

If ${}_np_{ij}$ is independent of n, we say that the transition probabilities are stationary and the matrix $P = (p_{ij})$ is called a *stochastic* matrix. It

follows that a stationary Markoff chain must have stationary transition
130 probabilities, but the converse is not necessarily true.

It is often useful to consider one sided chains, say for $n \geq 1$ and the behaviour of the chain then depends on the *initial state* or the initial probability distribution of x_1.

The theory of Markoff chains with a finite number of states can be treated completely (see for example J.L.Doob, Stochastic processes page 172). In the case of *stationary* transition probabilities, the matrix $({p^n}_{ij})$ defined by

$$p^1_{ij} = p_{ij},\ p^{n+1}_{ij} = \sum_k p^n_{ik}\ p_{kj}$$

satisfies $p^n_{ij} \geq 0$, $\sum_j p^n_{ij} = 1$ and gives the probability that a sequence with $x_1 = i$ will have $x_n = j$. The main problem is to determine the asymptotic behaviour of p^n_{ij}. The basic theorem is

Theorem 56 (For the proof see J.L.Doob, Stochastic Processes page 175)**.**

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} p^m_{ij} = q_{ij}$$

where $Q = (q_{ij})$ is a stochastic matrix and $QP = PQ = Q, Q^2 = Q$.

The general behaviour of p^n_{ij} can be described by dividing the states into transient states and disjoint ergodic sets of states. Almost all sequences have only a finite number of terms in any one of the transient states and almost all sequences for which x_n lies in an ergodic set will have all its subsequent terms in the same set.

A random function of a continuous variable t is called a Markoff if 131
, for $t_1 < t_2 \cdots < t_n < t$ and intervals (or Borel sets) $I_1, I_2, \ldots, I_n$, we have

$$P(x(t) \in I/x(t_1) \in I_1, x(t_2) \in I_2, ..., x(t_n) \in I_n) \\ = P(x(t) \in I/x(t_n) \in I_n).$$

Part of the theory is analogous to that of Markoff Chains, but the theory is less complete and satisfactory.

15. L_2-Processes

Theorem 57. *If $r(s,t)$ is the auto correlation function of an L_2 function,*

$$r(s,t) = \overline{r(t,s)}$$

and if (z_i) is any finite set of complex numbers, then

$$\sum_{i,j} r(t_i,t_j) z_i \overline{z}_j \geq 0.$$

The first part is trivial and the second part follows from the identity

$$\sum_{i,j} r(t_i,t_j) z_i \overline{z}_j = E(\ |(x(t_i) - m(t_i)) z_i|\ ^2) \geq 0.$$

Theorem 58 (For proof see J.L.Doob, Stochastic processes page 72)**.** *If $m(t), r(s,t)$ are given and $r(s,t)$ satisfies the conclusion of theorem 57, then there is a unique Gaussian function $x(t)$ for which*

$$E(x(t)) = m(t), E(x(s)\ \overline{x(t))} - m(s)\overline{m(t))} = r(s,t).$$

132

The uniqueness follows from the fact that a Gaussian process is determined by its first and second order moments given $r(s,t)$. Hence, if we are concerned only with properties depending on $r(s,t)$ and $m(t)$ we may suppose that all our processes are Gaussian.

Theorem 59. *In order that a centred L_2-process should be orthogonal, it is necessary and sufficient that*

$$E(|x(t) - x(s)|^2) = F(t) - F(s) (s < t) \tag{1}$$

where $F(S)$ is a non -decreasing function. In particular, if $x(t)$ is stationary L_2, then $E(|x(t) - x(s)|^2) = \sigma^2(t-s)(s < t)$ for some constant σ^2.

Proof. If $s < u < t$ the orthogonality condition implies that $E(|x(u) - x(s)|^2) + E(|x(t) - x(u)|^2) = E(|x(t) - x(s)|^2)$ which is sufficient to prove (1). The converse is trivial.

We write

$$dF = E(|dx|^2)$$

and for stationary functions,

$$\sigma^2 dt = E(|dx|^2).$$

We say that an L_2 - function is *continuous* at t if

$$\lim_{h \longrightarrow 0} E(|x(t+h) - x(t)|^2) = 0,$$

and that it is continuous if it is continuous for all t. Note that this does not imply that the individual $x(t)$ are continuous at t. □

Theorem 60 (Slutsky)**.** *In order that $x(t)$ be continuous at t, it is necessary and sufficient that $r(s,t)$ be continuous at $t = s$.* **133**

It is continuous (for all t) if $r(s,t)$ is continuous on the line t=s and then $r(s,t)$ is continuous in the whole plane.

Proof. The first part follows from the relations

$$\begin{aligned} & E(|x(t+h) - x(t)|^2) \\ & \quad = r(t+h, t+h) - r(t+h, t) - r(t, t+h) + r(t,t) \\ & \quad = o(1) \text{ as } h \longrightarrow 0 \text{ if } r(s,t) \text{ is continuous for} \\ & \qquad\qquad t = s; r(t+h, t+k) - r(t,t) \\ & \quad = E(x(t+h)\overline{x(t+k)} - x(t)\overline{x(t)}) \\ & \quad = E(x(t+h)\overline{x(t+k)} - \overline{x(t)}) + = ((x(t+h) - x(t))\overline{x(t)}) \\ & \quad = o(1) \text{ as } h, k \longrightarrow 0 \\ & \qquad \text{by the Schwartz inequality if } x(t) \text{ is continuous at } t. \end{aligned}$$

For the second part, we have

$$\begin{aligned} & r(s+h, t+k) - r(s,t) \\ & \quad = E(x(s+h)\overline{x(t+k)} - \overline{x(t)})) + E((x(s+h) - x(s))\overline{x(t)}) \\ & \quad = o(1) \text{as h, k} \longrightarrow 0 \text{ by Schwarz's inequality,} \\ & \qquad \text{if x(t) is continuous at t and s.} \end{aligned}$$

□

Theorem 61. *If $x(t)$ is continuous and stationary L_2, with $\rho(h) = r(s, s+h)$, then*

$$\rho(h) = \int_{-\infty}^{\infty} e^{i\lambda h} dS(\lambda)$$

where $S(\lambda)$ is non- decreasing and bounded.

Moreover,

$$S(\infty) - S(-\infty) = \rho(0) = E(|x(t)|^2) \text{ for all } t.$$

134 *Proof.* We have $\rho(-h) = \overline{\rho(h)}, \rho(h)$ is continuous at 0 and

$$\sum_{i,j} \rho(t_i - t_j) z_i \overline{z_j} \geq 0$$

for all complex z_i by Theorem 57 and the conclusion follows from Bochner's theorem (Loeve, Probability theory, p. 207-209, and Bochner Harmonic analysis and Probability, page 58).

The theorem for sequences is similar. □

Theorem 62. *If x_n is stationary L_2, with $\rho_n = E(x_m.\overline{x_{m+n}})$ then*

$$\rho_n = \int_{-\pi}^{\pi} e^{in\lambda} dS(\lambda)$$

where $S(\lambda)$ increases and $S(\pi) - S(-\pi) = \rho_o = E(| x_m |^2)$

We say that an L_2-random function $x(t)$ is differentiable at t with derivative $x'(t)$ (a random variable)if

$$E\left(| \frac{x(t+h) - x(t)}{h} - \dot{x}(t) |^2\right) \to 0 \text{ as } h \to 0$$

Theorem 63. *In order that $x(t)$ be differentiable at t it is necessary and sufficient that $\frac{\partial^2 r}{\partial s \partial t}$ exists when $t = s$. Moreover, if $x(t)$ is differentiable for all t, $\frac{\partial^2 r}{\partial s \partial t}$ exists on the whole plane.*

(The proof is similar to that of Theorem 60.)

Integration of $x(t)$ can be defined along the same lines.

We say that $x(t)$ is R-integrable in $a \leq t \leq b$ if $\sum_i x(t_i)\delta_i$ tends to
135 *a limit in L_2 for any sequence of sub-divisions of (a, b) into intervals of lengths δ_i containing points t_i respectively. The limit is denoted by $\int_a^b x(t)dt$.*

Theorem 64. *In order that $x(t)$ be R-integrable in $a \leq t \leq b$ it is necessary and sufficient that $\int_a^b \int_a^b r(s, t)dsdt$ exists as a Riemann integral.*

Riemann - Stieltjes integrals can be defined similarly.

The idea of integration with respect to a random function $Z(t)$ is deeper (see e.g. J.L.Doob, Stochastic processes, chap. IX §2). In the important cases, $Z(t)$ is orthogonal, and then it is easy to define the integral

$$\int_a^b \phi(t)dZ(t)$$

the result being a random variable. Similarly

$$\int_a^b \phi(s, t)dZ(t)$$

will be a random function of s under suitable integrability conditions.

The integral of a random function $x(t)$ with respect to a random function $Z(t)$ can also be defined (Doob, Chap. IX § 5).

The most important application is to the spectral representation of a stationary process.

Theorem 65 (Doob, page 527)**.** *A continuous stationary (L_2) function*
$x(t)$ can be represented in the form 136

$$x(t) = \int_{-\infty}^{\infty} e^{i\lambda t} dZ(\lambda)$$

where $Z(\lambda)$ has orthogonal increments and

$$E(| dZ |^2) = dS$$

where $S(\lambda)$ is the function defined in Theorem 61.

The formula gives the spectral decomposition of $x(t)$*.* $S(\lambda)$ *is its spectral distribution.*

The corresponding theorem for random sequence is

Theorem 66 (Doob, page 481)**.** *A stationary* (L_2) *sequence* $\{x_n\}$ *has spectral representation*

$$x_n = \int_{-\pi}^{\pi} e^{i\lambda n} dZ(\lambda)$$

where $Z(\lambda)$ *has orthogonal increments and*

$$E(| dZ |)^2 = dS,$$

$S(\lambda)$ *being defined by Theorem 62.*

Two or more random function $x_i(t)$ *are mutually orthogonal if* $E(x_i(t)\overline{x_j(s)}) = 0$ *for* $i \neq j$ *and* s, t*.*

Theorem 67. *Suppose that* $x(t)$ *is a continuous, stationary* (L_2) *process and that* $E_1, E_2 \ldots, E$ *are measurable, disjoint sets whose union is the whole real line. Then we can write*

$$x(t) = \sum_{i=1}^{\nu} x_i(t)$$

137 *where* $x_i(t)$ *are mutually orthogonal and*

$$x_i(t) = \int_{-\infty}^{\infty} e^{i\lambda t} dZ_i(\lambda) = \int_{E_i} e^{i\pi t} dZ(\lambda)$$

and $E(| dZ_i |)^2 = 0$ *outside* E_i*.*

The theorem for sequences is similar. In each case, a particularly important decomposition is that in which three sets E_i *are defined by the Lebesgue decomposition of* $S(\lambda)$ *into absolutely continuous, discontinuous and singular components. For the second component, the auto-correlation function* $\rho(n)$ *has the form*

$$\rho(h) = \sum_i d_i e^{ih\lambda i}$$

where d_i are the jumps of $S(\lambda)$ at the discontinuities λ_i, and is uniformly almost periodic.

We can define liner operations on stationary functions (Doob, page 534). In particular, if $k(s)$ of bounded variation in $(-\infty, \infty)$, the random function

$$y(t) = \int_{-\infty}^{\infty} x(t-s)dk(s)$$

can be defined and it is easy to show that y(t) has spectral representation

$$y(t) = \int_{-\infty}^{\infty} e^{i\lambda t} K(\lambda) dZ(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda t} dz_1(\lambda)$$

where $$K(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda s} dk(s), E(\mid dZ_1(\lambda)\mid^2) = (K(\lambda))^2 dS(\lambda).$$

138

If $k(s) = 0$ *for* $s < \tau, \tau > 0$ *we have*

$$y(t+\tau) = \int_0^{\infty} x(t-s)dk(s-\tau)$$

which depends only on the "part" of the function $x(t)$ "before time t". The linear prediction problem (Wiener) is to determine $k(s)$ so as to minimise (in some sense) the difference between $y(t)$ and $x(t)$. In so far as this difference can be made small, we can regard $y(t+\tau)$ as a prediction of the value of $x(s)$ at time $t+\tau$ based on our knowledge of its behaviour before t.

16. Ergodic Properties

We state first the two basic forms of the ergodic theorem.

Theorem 68 (G.D Birkhoff, 1932)**.** *Suppose that for $\lambda \geq 0, T^{\lambda}$ is a measure preserving (1-1) mapping of a measure space Ω of measure 1 onto itself and that $T^0 = I$, $T^{\lambda+\mu} = T^{\lambda} \circ T^{\mu}$. Suppose that $f(\omega) \in$*

$L(\Omega)$and that $f(T^{\lambda}\Omega)$ is a measurable function of (λ,ω) in the product space $R\times\Omega$. Then

$$f^*(\omega) = \lim_{\lambda\to\infty} \frac{1}{\lambda}\int_0^{\lambda} f(T^{\Lambda}\omega)d\lambda$$

exists for almost all ω, $f^*(\omega)\in L(\Omega)$ and

$$\int_{\Omega} f^*(\omega)d\omega = \int_{\Omega} f(\omega)d\omega$$

139 *Moreover, if Ω has no subspace of measure > 0 and < 1 invariant under all T^{λ}*

$$f^*(\omega) = \int_{\Omega} f(\omega)d\omega \text{ for almost all } \omega$$

There is a corresponding discrete ergodic theorem for transformations $T^n = (T)^n$ where n is an integer, the conclusion then being that

$$f^*(\omega) = \lim_{n\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(T^n\omega)$$

140 *exists for almost all ω. In this case, however, the memorability condition on $f(T^{\lambda}\omega)$ may be dispensed with.*

Theorem 69 (Von Neumann)**.** *Suppose that the conditions of Theorem 68 hold and that $f(\omega)\in L_2(\Omega)$. Then*

$$\int_{\Omega}\left|\frac{1}{\Lambda}\int_0^{\lambda} f(T^{\lambda}\omega)d\lambda - f^*(\omega)\right|^2 d\omega \to 0 \text{ as } \wedge \to \infty.$$

For proofs see Doob page 465, 515 or P.R. Halmos, Lectures on Ergodic Theory, The math.Soc. of Japan, pages 16,18. The simplest proof is due to F.Riesz (Comm. Math.Helv. 17 (1945)221-239).

Theorems 68 is much than Theorem 69.

The applications to random functions are as follows

Theorem 70. *Suppose that $x(t)$ is a strictly stationary random function and that $x(\omega, t) \in L(\Omega)$ for each t, with $\int_\Omega x(\omega, t) d\omega = E(x(t)) = m$. Then*

$$\lim_{\Lambda \to \infty} \frac{1}{\Lambda} \int_0^\Lambda x(\omega, t) dt = x^*(\omega)$$

exists for almost all ω If $x(t)$ is an L_2- function we have also convergence in mean .

This follows at once from Theorem 68, 69 if we define

$$f(\omega) = x(\omega, 0), T^\lambda(x(t)) = x(t + \lambda).$$

Corollary. *If a is real*

$$\lim_{\Lambda \to \infty} \frac{1}{\Lambda} \int_0^\Lambda x(\omega, t) e^{iat} dt = x^*(\omega, a)$$

exists for almost all ω.

Theorem 70 is a form of the strong law of large number for random functions. There is an analogue for sequences.

A particularly important case arises if the translation operation $x(t) \to x(t + \lambda)$ has no invariant subset whose measure is >0 and < 1. In this case we have

$$\lim_{\Lambda \to \infty} \frac{1}{\Lambda} \int_0^\Lambda x(\omega, t) dt = \int_\Omega x(\omega, t) d\omega = m$$

for almost all ω. In other words almost all functions have limiting "time averages" equal to the mean of the values of the function at any fixed time.

17. Random function with independent increments

The basic condition is that if $t_1 < t_2 < t_3$, then $x(t_2) - x(t_1)$ and $x(t_3) - x(t_2)$ are independent (see page 114) so that the distribution function of

$x(t_3) - x(t_1)$ is the convolution of those of $x(t_3) - x(t_1)$ and $x(t_3) - x(t_2)$. We are generally interested only in the increments, and it is convenient to consider the behaviour of the function from some base point, say **141**
0, modify the function by subtracting a random variable so as to make $x(0) = 0$, $x(t) = x(t) - x(0)$. Then if F_{t_1,t_2} is the distribution function for the increment $x(t_2) - x(t_1)$ we have,

$$F_{t_1,t_3}(x) = F_{t_1,t_2} * F_{t_2,t_3}(x).$$

We get the stationary case if $F_{t_1,t_2}(x)$ depends only on $t_2 - t_1$. (This by itself is not enough, but together with independence, the condition is *sufficient* for stationary) If we put

$$F_t(x) = F_{o.t}(x),$$

we have in this case

$$F_{t_1+t_2}(x) = F_{t_1} * F_{t_2}(x),$$

for all $t_1, t_2 > 0$.

If $x(t)$ is also an L_2 function with $x(0) = 0$, it follows that

$$E(|x(t_1 + t_2)|^2) = E(|x(t_1)|^2) + E(|x(t_2)|^2)$$

so that $$E(|x(t)|^2) = t\sigma^2$$

where $$\sigma^2 = E(|x(1)|^2)$$

Theorem 71. *If $x(t)$ is stationary with independent increments, its distribution function $F_t(x)$ infinitely divisible and its characteristic function $\varphi_t(u)$ has the form $e^{t\psi(u)}$, where*

$$\psi(u) = iau + \int_{-\infty}^{\infty} \left[e^{itx} - 1 - \frac{iux}{1+x^2}\right] \frac{1+x^2}{x^2} dG(x)$$

142 *$G(x)$ being non-decreasing and bounded,*

Proof. The distribution function is obviously infinitely divisible for every t and it follows from the stationary property that

$$\varphi_{t_1+t_2}(u) = \varphi_{t_1}(u)\varphi_{t_2}(u)$$

so that $\varphi_t(u) = e^{t\psi(u)}$ for some $\psi(u)$, which must have the $K - L$ form which is seen by putting $t = 1$ and using Theorem 37. □

Conversely, we have also the

Theorem 72. *Any function $\varphi_t(u)$ of this form is the characteristic function of a stationary random function with independent increments.*

Proof. We observe that the conditions on $F_t(x)$ gives us a system of joint distributions over finite sets of points t_i which is consistent in the sense of Theorem 44 and the random function defined by the Kolmogoroff measure in Theorem 44 has the required properties. □

Example 1 (Brownian motion : Wiener)**.** The increments all have normal distributions, so that

$$F_t(x) = \frac{1}{\sigma\sqrt{2\pi t}} e^{-x^2/2t\sigma^2}$$

Example 2 (Poisson)**.** The increments $x(s+t)-x(s)$ have integral values $\nu \geq 0$ with probabilities $e^{-ct}\frac{(ct)^\nu}{\nu!}$

Both are L_2- Processes. **143**

Theorem 73. *Almost all functions $x(t)$ defined by the Kolmogoroff measure defined by the Wiener function, or any extension of it, are everywhere non-differentiable. In fact, almost all functions fail to satisfy a Lipschitz condition of order $\alpha(x(t-h) - x(t) = 0(|\ h\ |^\alpha))$ if $\alpha > \frac{1}{2}$ and are not of bounded variation.*

Theorem 74. *The Kolmogoroff measure defined by the Wiener function can be extended so that almost all functions $x(t)$ Satisfy a Lipschitz condition of any order $\alpha < \frac{1}{2}$ at every point, and are therefore continuous at every point.*

For proofs, see Doob, pages 392-396 and for the notion of extension, pages 50-71.

Theorem 75. *The K-measure defined by the Poisson function can be extended so that almost functions $x(t)$ are step functions with a finite number of positive integral value in any finite interval.*

The probability that $x(t)$ will be constant in an interval of length t is e^{-ct}.

18. Doob Separability and extension theory

The K-measure is usually not extensive enough to give probabilities to important properties of the functions $x(t)$, e.g. continuity etc.

Doob's solution is to show that a certain subject $\Omega_\circ$ of Ω has outer
144 K-measure 1, $\mu(\Omega_\circ) = 1$. Then, if X_1 is any K-measurable set, Doob defines

$$\mu\star(X) = \mu(X_1) \text{when } X = \Omega_o X_1$$

and shows that $\mu^\star$ is completely additive and defines a probability measure in a Borel system *containing* Ω_0, and $\mu^\star(\Omega_0) = 1$.

Doob now defines a *quasi-separable* K-measure as one for which there is a subset Ω_0 of outer K-measure 1 and a countable set R_0 of real numbers with the property that

$$\sup_{t\epsilon I} x(t) = \sup_{t\epsilon I.R_\circ} x(t)$$
$$\inf_{t\epsilon I} x(t) = \inf_{t\epsilon I.R_\circ} x(t) \qquad (\alpha)$$

for every $x(t) \in \Omega_\circ$ and every open interval I.

If the K-measure has this property, it can be extended to a measure so that almost all functions $x(t)$ have the property (α).

All conditions of continuity, differentiability and related concepts can be expressed then in terms of the countable set R_0 and the sets of functions having the corresponding property then become measurable. Thus, in the proofs of Theorem 74 we have only to show that the set of functions having the required property (of continuity or Lipschitz condition) has outer measure 1 with respect to the basic Wiener measure.

For Theorem 73, there is no need to extended the measure, for if
145 the set of functions $x(t)$ which are differentiable at least at one point has measure zero, with respect to Wiener measure, it has measure zero with respect to any extension of Wiener measure.

For a fuller account, see Doob, Probability in Function Space, Bull. Amer. Math. Soc. Vol. 53 (1947),15-30.