Chapter I Standard Borel Spaces

1.1 σ-algebras

Let S be a space. A class $\mathcal{S}$ of subsets of S is called a σ-algebra on S if (1) $S \in \mathcal{S}$, (2) $A \in \mathcal{S} \Longrightarrow A^c \in \mathcal{S}$ and (3) $A_n \in \mathcal{S}$, $n = 1,2,\ldots \Longrightarrow \bigcup_n A_n \in \mathcal{S}$. Let $\mathcal{C}$ be an arbitrary class of subsets of S. Then the intersection of all σ-algebras on S that include $\mathcal{C}$ is also a σ-algebra on S. It is called the σ-algebra generated by $\mathcal{C}$, $\sigma(\mathcal{C})$ in notation.

Let $\mathcal{S}$ be a σ-algebra on S and T a subset of S. Then $T \cap \mathcal{S} = \{T \cap A: A \in \mathcal{S}\}$ is a σ-algebra on T, called the trace of $\mathcal{S}$ on T.

Let $\mathcal{S}_\lambda$ be a σ-algebra on S_λ for $\lambda \in \Lambda$. Let S be the (Cartesian) product of S_λ, $\lambda \in \Lambda$, i.e. $S = \Pi_\lambda S_\lambda$. Let π_λ be the λ-projection from S onto S_λ. The σ-algebra on S generated by the sets: $\pi_\lambda^{-1}(A_\lambda)$, $\lambda \in \Lambda$, $A_\lambda \in \mathcal{S}_\lambda$, is called the product σ-algebra of $\mathcal{S}_\lambda$, $\lambda \in \Lambda$, $\otimes_\lambda \mathcal{S}_\lambda$ in notation.

Suppose that S is a topological space with topology τ. The The σ-algebra generated by all τ-open subsets of S is called the topological σ-algebra. $\mathcal{B}_\tau(S)$ in notation. The suffix τ is often omitted if there is no possibility of confusion. If we have two topologies τ and σ on S and if τ is stronger than σ, then every σ-open set is also τ-open and therefore we have $\mathcal{B}_\tau(S) \supset \mathcal{B}_\sigma(S)$ in general. It often happens that two different topologies may induce the same topological σ-algebra. (See Example 1 at the end of this section.)

Let $S = (S,\tau)$ be a topological space. Then every subset T of S is regarded as a topological space with the relative topology $\tau|T$ in which a set $G \subset T$ is called open in T if it is the intersection of T and an τ-open subset of S. It is easy to verify

(1) $$\mathcal{B}_{\tau|T}(T) = T \cap \mathcal{B}(S).$$

Let $S_\lambda = (S_\lambda, \tau_\lambda)$ be a topological space for $\lambda \in \Lambda$. Then the product space $S = \Pi_\lambda S_\lambda$ is also a topological space with the product topology τ.

Then we have

(2) $$\mathcal{B}_\tau(S) \supset \bigotimes_\lambda \mathcal{B}_{\tau_\lambda}(S_\lambda)$$

and they are not equal in general. (See Example 2 below.)

<u>Theorem 1</u>. If Λ is countable and if the topology τ_λ in S_λ has a countable open base for each $\lambda \in \Lambda$, then the product topology in $\Pi_\lambda S_\lambda$ also has a countable open base and the two σ-algebras in (2) are the same.

<u>Proof</u> Obvious by the definitions.

Let $\mathcal{S}$ and $\mathcal{T}$ be σ-algebras on S and T respectively. A map $f: S \to T$ is called <u>measurable</u> $\mathcal{S}/\mathcal{T}$ if $f^{-1}(\mathcal{T}) \subset \mathcal{S}$ i.e. $f^{-1}(B) \in \mathcal{S}$ for $B \in \mathcal{S}$. If $f: S \to T$ is measurable $\mathcal{S}/\mathcal{T}$ and if $g: T \to U$ is measurable $\mathcal{T}/\mathcal{U}$, then the composite map $g \circ f$: $S \to U$ is measurable $\mathcal{S}/\mathcal{U}$. If T is a topological space, then we often use the term "measurable $\mathcal{S}$" instead of measurable $\mathcal{S}/\mathcal{B}(T)$.

<u>Theorem 2</u> Let $\mathcal{S}$ be a σ-algebra on S and T a metric space with metric ρ. Let $\mathcal{B}(T)$ denote the topological σ-algebra on T.

(i) The limit of a sequence of functions: $S \to T$ measurable $\mathcal{S}/\mathcal{B}(T)$ is also measurable $\mathcal{S}/\mathcal{B}(T)$.

(ii) If T is a separable metric space, then every function: $S \to T$ measurable $\mathcal{S}/\mathcal{B}(T)$ can be expressed as the uniform limit of a sequence of functions: $S \to T$ measurable $\mathcal{S}/\mathcal{B}(T)$ and taking a countable number of values.

Proof

(i) Let G be any open set. Take a sequence of open sets $\{G_p\}$ such that $G_1 \subset \overline{G}_1 \subset G_2 \subset \overline{G}_2 \subset G_3 \subset \dots \to G$, for example

$$G_p = \{ x \in G: \rho(x,y) > 1/p \text{ for } y \in G^c \}.$$

It is obvious that G_{p+1} includes the closure $\overline{G}_p$ of G_p. Therefore

$$f^{-1}(G) = \bigcup_p f^{-1}(G_p) = \bigcup_p \bigcup_k \bigcap_{n \geq k} f_n^{-1}(G_p) \in \mathcal{S}.$$

(ii) Let $f: S \to T$ be measurable $\mathcal{S}/\mathcal{B}(T)$ Take a countable dense subset $\{a_m\}$ in T and let U_{mn} denote the ball $\{ y \in T: \rho(y, a_m) < 1/n\}$ for m, n = 1, 2, ... Set

$$V_{mn} = U_{mn} - \bigcup_{k<m} U_{kn} \text{ and } A_{mn} = f^{-1}(V_{mn}).$$

Then

$$A_{mn} \in \mathcal{S} \text{ and } S = \bigcup_m A_{mn} \text{ (disjoint union).}$$

Define $f_n: S \to T$ by

$$f_n(x) = a_m \text{ for } x \in A_{mn}, \ m = 1, 2, \dots .$$

Then f_n takes only values in $\{a_1, a_2, \dots\}$ and

$$\rho(f_n(x), f(x)) = \rho(a_m, f(x)) < 1/n \text{ for } x \in A_{mn}$$

This completes the proof.

Let $\mathcal{S}_i$ be a σ-algebra on S_i for $i = 1,2,3$. For $A \subset S_1 \times S_2$ we consider the <u>sections</u>

$$A(x_1) = \{x_2 \in S_2: \ (x_1, x_2) \in A\} \quad , \text{ for fixed } x_1 \in S_1$$

and

$$A(x_2) = \{x_1 \in S_1: \ (x_1, x_2) \in A\} \quad , \text{ for fixed } x_2 \in S_2,$$

Similarly for $f: S_1 \times S_2 \to S_3$ we consider the <u>sections</u>

$$f_{x_1}: \ S_2 \to S_3 \quad , \ f_{x_1}(x_2) \equiv f(x_1, x_2) \text{ for fixed } x_1 \in S_1,$$

and

$$f_{x_2}: \ S_1 \to S_3 \quad , \ f_{x_2}(x_1) = f(x_1, x_2) \text{ for fixed } x_2 \in S_2$$

Then we have the following

<u>Theorem 3</u>

(i) If $A \in \mathcal{S}_1 \otimes \mathcal{S}_2$, then $A(x_1) \in \mathcal{S}_2$ for fixed $x_1 \in S_1$ and $A(x_2) \in \mathcal{S}_1$ for fixed $x_2 \in S_2$.

(ii) If $f: S_1 \times S_2 \to S_3$ is measurable $\mathcal{S}_1 \otimes \mathcal{S}_2/\mathcal{S}_3$, then $f_{x_1}: S_2 \to S_3$ is measurable $\mathcal{S}_2/\mathcal{S}_3$ for fixed $x_2 \in S_2$ and $f_{x_2}: S_1 \to S_3$ is measurable $\mathcal{S}_1/\mathcal{S}_3$ for fixed $x_2 \in S_2$.

<u>Proof</u> (i)If $A = A_1 \times A_2$, $A_i \in \mathcal{S}_i$, then

$$\begin{aligned} A(x_1) &= A_2 \quad \text{if } x_i \in A_1, \\ &= \emptyset \quad \text{if } x_i \notin A_1. \end{aligned}$$

and so $A(x_1) \in \mathcal{S}_2$ for $x_1 \in S_1$. Since the class of all sets $A \subset S_1 \times S_2$ for which $A(x_1) \in \mathcal{S}_2$ for every $x_1 \in S_1$ is a σ-algebra on $S_1 \times S_2$, this class includes the σ-algebra $\mathcal{S}_1 \otimes \mathcal{S}_2$. This proves the first part of (i). Similarly for the second part.

(ii) This is an immediate result from (i) by virtue of the obvious relations

$$f_{x_1}^{-1}(B) = (f^{-1}(B))(x_1) \text{ and } f_{x_2}^{-1}(B) = (f^{-1}(B))(x_2).$$

The converse of Theorem 3 is not true in general. A set or a function in the product space is not always measurable even if its sections are measurable. The following theorem is useful in this connection.

Theorem 4. Suppose that

(1) $\mathcal{S}_1$ is a σ-algebra on S_1,

(2) S_2 is a separable metric space,

(3) S_3 is a metric space.

and

(4) f is a map: $S_1 \times S_2 \to S_3$.

If the section $f_{x_1}: S_2 \to S_3$ is continuous and if the section $f_{x_2}: S_1 \to S_3$ is measurable $\mathcal{S}_1/\mathcal{B}(S_3)$, then f is measurable $\mathcal{S}_1 \otimes \mathcal{B}(S_2)/\mathcal{B}(S_3)$.

Proof. Since the identity map $I: S_2 \to S_2$ is measurable $\mathcal{B}(S_2)/\mathcal{B}(S_2)$ we can use Theorem 2(ii) to obtain a sequence of $I_n: S_2 \to S_2$ measurable $\mathcal{B}(S_2)/\mathcal{B}(S_2)$ such that the set C_n of values of I_n is countable for each n and that

$$I_n(x_2) \to x_2 (n \to \infty) \quad \text{for every} \quad x_2 \in S_2.$$

Set $g_n(x_1,x_2) = f(x_1, I_n(x_2))$. Then

$$g_n(x_1,x_2) \to f(x_1,x_2) \quad (n \to \infty).$$

and

$$g_n^{-1}(A) = \bigcup_{c \in C_n} f_c^{-1}(A) \times \{c\}$$

Therefore g_n is measurable $\mathcal{S}_1 \otimes \mathcal{B}(S_2)/\mathcal{B}(S_3)$ and so is f.

Theorem 5. If $f_\lambda: S \to T_\lambda$ is measurable $\mathcal{S}/\mathcal{T}_\lambda$ for every $\lambda \in \Lambda$. Then the map $f: S \to \prod_\lambda T_\lambda$ defined by $f(x) = (f_\lambda(x), \lambda \in \Lambda)$ is measurable $\mathcal{S}/\otimes_\lambda \mathcal{T}_\lambda$.

Proof. Write T and $\mathcal{T}$ respectively for $\Pi_\lambda T_\lambda$ and $\otimes_\lambda \mathcal{T}_\lambda$ and let $\pi_\lambda: T \to T_\lambda$ be the λ-projection. Then

$$f^{-1}(\pi_\lambda^{-1}(B_\lambda)) = (\pi_\lambda \circ f)^{-1}(B_\lambda) = f_\lambda^{-1}(B_\lambda) \in \mathcal{S}$$

for $B_\lambda \in \mathcal{T}_\lambda$. This implies

$$\Pi_\lambda^{-1}(B_\lambda) \in \mathcal{T}^1 \equiv \{B \in T: f^{-1}(B) \in \mathcal{S}\}.$$

But $\mathcal{T}^1$ is clearly a σ-algebra on T. Therefore $\mathcal{T}^1 \supset \mathcal{T}$, i.e. f is measurable $\mathcal{S}/\mathcal{T}$.

Theorem 6. If $f_\lambda: S_\lambda \to T_\lambda$ is measurable $\mathcal{S}_\lambda/\mathcal{T}_\lambda$ for every $\lambda \in \Lambda$, then the map $f: \Pi_\lambda S_\lambda \to \Pi_\lambda T_\lambda$ defined by

$$f(x_\lambda, \lambda \in \Lambda) = (f_\lambda(x_\lambda) \lambda \in \Lambda)$$

is measurable $\otimes_\lambda \mathcal{S}_\lambda/\otimes_\lambda \mathcal{T}_\lambda$.

Proof. Write S and $\mathcal{S}$ for $\Pi_\lambda S_\lambda$ and $\otimes_\lambda \mathcal{S}_\lambda$ and let $\pi_\lambda: S \to S_\lambda$ be the λ-projection. Then

$$f(x) = (f_\lambda \circ \pi_\lambda (x), \lambda \in \Lambda), \ x \in S.$$

Since π_λ is measurable $\mathcal{S}/\mathcal{S}_\lambda$ and since f_λ is measurable $\mathcal{S}_\lambda/\mathcal{T}_\lambda$, $f_\lambda \circ \pi_\lambda$ is measurable $\mathcal{S}/\mathcal{T}_\lambda$. Now use the previous theorem.

Example 1. Let H be a separable Hilbert space and let σ and ω be the norm topology and the weak topology respectively. Since σ is stronger than ω, we have $\mathcal{B}_\sigma(H) \supset \mathcal{B}_\omega(H)$. Every σ-open set is a countable union of σ-closed balls and every σ-closed ball is ω-closed. Therefore every σ-open set belongs to $\mathcal{B}_\omega(H)$. This implies $\mathcal{B}_\sigma(H) \subset \mathcal{B}_\omega(H)$. Therefore we have $\mathcal{B}_\sigma(H) = \mathcal{B}_\omega(H)$, although σ is strictly stronger than ω.

Example 2. If Λ is uncountable, then

$$\mathcal{B}(R^\Lambda) \supsetneqq \mathcal{B}(R)^{\otimes\Lambda}$$

(If $S_\lambda = T$ and $\mathcal{S}_\lambda = \mathcal{T}$ for $\lambda \in \Lambda$, then we write T^Λ and $\mathcal{T}^{\otimes\Lambda}$ respectively for $\Pi_\lambda S_\lambda$, and $\otimes_\lambda \mathcal{S}_\lambda$.)

To prove this, we introduce the notion "countably determined." A subset A of R^Λ is called countably determined if we have a countable subset $M = M_A$ of Λ such that

$$(x_\lambda) \in A \text{ and } x_\lambda = y_\lambda \text{ for } \lambda \in M \Longrightarrow (y_\lambda) \in A.$$

The class $\mathcal{C}$ of all countably determined subsets of R^Λ is a σ-algebra on R^Λ. Since the set $\Pi_\lambda^{-1}(E_\lambda)$ is countably determined for every $\lambda \in \Lambda$ and $E_\lambda \in \mathcal{S}_\lambda$, we have $\mathcal{C} \supset \mathcal{B}(R^1)^{\otimes\Lambda}$. Therefore every set $\in \mathcal{B}(R)^\Lambda$ is countably determined. Since Λ is uncountable, every singleton (= single point set) is not countably determined but belongs to $\mathcal{B}(R^\Lambda)$ as a closed subset of R^Λ.

Example 3. For $\Lambda = \{1,2,3,\ldots n\}$ or $\{1,2,3\ldots\}$ we write R^n or R^∞ for R^Λ. By Theorem 1 we have $\mathcal{B}(R^\Lambda) = \mathcal{B}(R)^{\otimes\Lambda}$ in this case. We write $\mathcal{B}^n$ or $\mathcal{B}^\infty$ for this σ-algebra according as $\Lambda=\{1,2,\ldots n\}$ or $\{1,2,3,\ldots\}$.

<u>Example 4</u>. The two point set $\{0,1\}$ is a topological space with discrete topology. $\{0,1\}^n, n=1,2,\ldots,\infty$ are defined as above. The product σ-algebra and the topological σ-algebra for the product topology are the same. $\{0,1\}^\infty$ is called the <u>coin tossing space</u>, Γ in notation, because it is the sample space of coin tossing game.

1.2 Borel spaces

A space S endowed with a σ-algebra $\mathcal{S}$ on S is called a Borel space $(S, \mathcal{S})$. S is called the base space of $(S, \mathcal{S})$ and $\mathcal{S}$ the Borel structure in $(S, \mathcal{S})$. If there is no possibility of confusion, we will simply write S for $(S, \mathcal{S})$. A subset A of S is called a Borel subset if $A \in \mathcal{S}$.

To get rid of the trouble of assigning the Borel structure all the time, we will make the following convention.

(a) A topological space is regarded as a Borel space endowed with the topological σ-algebra.

(b) A subspace of a Borel space is regarded as a Borel space endowed with the trace σ-algebra.

(c) The product of Borel spaces is regarded as a Borel space endowed with the product σ-algebra.

Because of the relation (1) in 1.1 there is no contradiction between (a) and (b), but (b) and (c) may not be consistent. Therefore in the case of the topological product we should specify the Borel structure we are referring to, unless the two σ-algebras in (2) are known to be equal. (See 1.1 Theorem 1 and Example 3 .)

Let f be a map from a Borel space $(S, \mathcal{S})$ into another Borel space $(T, \mathcal{T})$. f is called Borel measurable if it is measurable $\mathcal{S}/\mathcal{T}$. A 1-1 map f from $(S, \mathcal{S})$ onto $(T, \mathcal{T})$ is called a Borel isomorphic map if both f and f^{-1} are Borel measurable, i.e. if $f(\mathcal{S}) = \mathcal{T}$. If there is such f, then $(T, \mathcal{T})$ is said to be Borel isomorphic with $(S, \mathcal{S})$. Borel isomorphism is an equivalence relation.

Theorem 1. If T is Borel isomorphic with S, then every Borel subset F of T is Borel isomorphic with a Borel subset of S.

Proof. Let f be a Borel isomorphic map from S onto T. Then $E = f^{-1}(F)$ is Borel isomorphic with F.

Theorem 2. If T_λ is Borel isomorphic with S_λ for every $\lambda \in \Lambda$, then $\Pi_\lambda T_\lambda$ is Borel isomorphic with $\Pi_\lambda S_\lambda$.

Proof. Let $f_\lambda\colon S_\lambda \to T_\lambda$ be a Borel isomorphic map for $\lambda \in \Lambda$. The map $f\colon \Pi_\lambda T_\lambda \to \Pi_\lambda S_\lambda$ defined by $f(x_\lambda, \lambda \in \Lambda) = (f_\lambda(x_\lambda), \lambda \in \Lambda)$ is Borel isomorphic, as can be easily checked.

Theorem 3. Every homeomorphic map is Borel isomorphic. Therefore two homeomorphic topological spaces are Borel isomorphic.

Proof. Recall the fact that a homorphic map $f\colon S \to T$ carries the open sets in S onto the open sets in T.

Theorem 4. Let S and T be Borel spaces and suppose that $S = \bigcup_n S_n$ and $T = \bigcup_n T_n$ are disjoint countable decompositions into Borel subsets. If T_n is a Borel isomorphic with S_n for every n, then T is Borel isomorphic with S.

Proof. Let f_n be a Borel isomorphic map from S_n onto T_n. Define $f\colon S \to T$ by $f = f_n$ on S_n $n = 1, 2, \ldots$. Then f is a Borel isomorphic map. To prove this, take an arbitrary Borel subset B of S. Then $B = \bigcup_n S_n \cap B$ (disjoint union) and so

$$f(B) = \bigcup_n f(S_n \cap B) = \bigcup_n f_n(S_n \cap B).$$

Since $S_n \cap B$ is a Borel subset of S_n, $f_n(S_n \cap B)$ is a Borel subset of T_n and so a Borel subset of T because T_n is a Borel subset of T. Therefore f(B) is Borel subset of T. This proves that f^{-1} is Borel measurable. Applying the same argument for f^{-1}, we have that $f = (f^{-1})^{-1}$ is also Borel measurable.

<u>Theorem 5</u>. Let $(S, \mathcal{B})$ be a Borel space.

(1) If Λ and M have the same cardinal number, then $(S^\Lambda, \mathcal{B}^{\otimes\Lambda})$ is Borel isomorphic with $(S^M, \mathcal{B}^{\otimes M})$.

(2) If the cardinal number of Λ is the sum of the cardinal numbers of Λ_α, $\alpha \in A$, then $(S^\Lambda, \mathcal{B}^\Lambda)$ is Borel isomorphic with $(\Pi_\alpha S^{\Lambda_\alpha}, \otimes_\alpha \mathcal{B}^{\Lambda_\alpha})$.

<u>Proof</u>. The same as the proof of similar facts on the topological product.

<u>Example</u>. $\Gamma^n (n=1,2,\ldots\infty)$ is Borel isomorphic (in fact homeomorphic) with Γ, where $\Gamma = \{0,1\}^\infty$ (the coin tossing space).

1.3 Standard Borel spaces

A Borel space $(S, \mathcal{S})$ is called standard if it is Borel isomorphic with a Borel subset of R^1. The notion of standard Borel spaces was introduced by G. MacKey [1] in connection with group representations.

Theorem 1. Every Borel space isomorphic with a standard Borel space is standard.

Proof. Obvious by transitivity of Borel isomorphism.

Theorem 2. Every Borel subset of a standard Borel space is a standard Borel space.

Proof. Obvious by 1.2 Theorem 1.

Theorem 3. A Borel space is standard if and only if it is Borel isomorphic with a Borel subset of the coin tossing space Γ.

Proof. By 1.2 Theorem 1 it is enough to prove that Γ is Borel isomorphic with R^1. Since R^1 is homeomorphic and so Borel isomorphic with $I = (0,1)$, it is enough to prove that $I = (0,1)$ is Borel isomorphic with Γ. Consider a map f from Γ onto $\bar{I}=[0,1]$:

$$f(x_1, x_2, \ldots) = \sum_{n=1}^{\infty} \frac{x_n}{2^n} \qquad (x_n = 0 \text{ or } 1).$$

Since $U_n(x) = \{y = (y_1, y_2, \ldots): y_i = x_i,\ i = 1,2,\ldots,n\}$, $n = 1,2,\ldots$, form a complete neighborhood system at $x = (x_1, x_2, \ldots)$, f is continuous. Let Γ_0 be the set of all $x = (x_1, x_2, \ldots)$ for which either $x_n = 0$ eventually or $x_n = 1$ eventually. Set $\Gamma_1 = \Gamma - \Gamma_0$ and define a map $f_1: \Gamma_1 \to I_1 = f(\Gamma_1) \subset I$ by

$$f_1(x) = f(x), \quad x \in \Gamma_1$$

Then f_1 is continuous, 1-1 and onto. For $x \in \Gamma_1$ and every n, we can find $p = p(x,n) > n$ and $q = q(x,n) > n$ such that $x_p = 0$ and $x_q = 1$. Set $r = r(x,n) = p \vee q$. Then it is easy to check that

$$|f(y) - f(x)| < 2^{-r} \implies y_i = x_i,\ i = 1,2,\dots n.$$

This implies that $f_1^{-1}: I_1 \rightarrow \Gamma_1$ is continuous. Therefore Γ_1 is homeomorphic and so Borel isomorphic with I_1. It is obvious that Γ_0 and $I_0 = I - I_1$ are both countably infinite, so that they are Borel isomorphic. Using 1.2 Theorem 4, we have that $I = I_1 \cup I_0$ is Borel isomorphic with $\Gamma = \Gamma_1 \cup \Gamma_0$.

Theorem 4. The product of a countable number of standard Borel spaces is a standard Borel space.

Proof. We will discuss the countably infinite product case, since the finite product case can be treated similarly. Let S_n, $n = 1,2,\dots$ be standard Borel spaces. Then each S_n is Borel isomorphic with a Borel subset E_n of Γ by Theorem 3. Therefore the product $S = \prod_n S_n$ is Borel isomorphic with $E = \prod_n E_n$ by 1.2 Theorem 2. But

$$E = \bigcap_n \pi_n^{-1}(E_n),$$

where $\pi_n: \Gamma^\infty \to \Gamma$ is the projection that carries $x \in \Gamma^\infty$ to its n-th component. Therefore E is a Borel subset of Γ^∞. Since Γ^∞ is Borel isomorphic with Γ (1.2 Example), Γ^∞ is a standard Borel space by Theorem 3. Therefore E is a standard Borel space by Theorem 2 and so is S by Theorem 1.

Theorem 5. Let S be a Borel space and S_n a Borel subset of S for $n = 1,2,\dots$ (finite or countably infinite) such that $S = \bigcup_n S_n$.

If each S_n is a standard Borel space, then S is also standard.

Proof. Set $T_n = S_n - \cup_{k=1}^{n-1} S_k$ $(\cup_{k=1}^{0} S_k = \emptyset)$. Since $T_n = S_n \cap (S - \cup_{k=1}^{n-1} S_k)$ and since $S - \cup_{k=1}^{n-1} S_k$ is a Borel subset of S, T_n is a Borel subset of S_n. As S_n is a standard Borel space, so is T_n by Theorem 2. Since R^1 is homeomorphic and so Borel isomorphic with $I_n = (n, n+1)$ and since T_n is Borel isomorphic with a Borel subset of R^1, T_n is Borel isomorphic with a Borel subset E_n of I_n by Theorem 1. Since $\{T_n\}$ and $\{E_n\}$ are both disjoint, we can apply 1.2 Theorem 4 to see that $S = \cup T_n$ and $E = \cup E_n$ are Borel isomorphic. As each E_n is obviously a Borel subset of R^1, so is E. This proves that S is standard.

Theorem 6. Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be Borel spaces and suppose that $(T, \mathcal{T})$ is standard. If $f: S \to T$ and $g: S \to T$ are both Borel measurable, then $\{x \in S: f(x) = g(x)\} \in \mathcal{S}$.

Proof. We can assume with no loss of generality that T is a Borel subset of R^1 and $\mathcal{T} = T \cap \mathcal{B}(R^1)$. Then

$$\begin{aligned}
&\{x \in S: f(x) = g(x)\} \\
&= \bigcap_n \bigcup_k \{x \in S: \tfrac{k}{n} \leq f(x) < \tfrac{k+1}{n}, \tfrac{k}{n} \leq g(x) < \tfrac{k+1}{n}\} \\
&= \bigcap_n \bigcup_k f^{-1}([\tfrac{k}{n}, \tfrac{k+1}{n}) \cap T) \cap g^{-1}([\tfrac{k}{n}, \tfrac{k+1}{n}) \cap T) \in \mathcal{S}.
\end{aligned}$$

Examples. The following spaces and their Borel subsets are all standard.

1. A countable set (with the discrete topology)
2. The n-space R^n
3. The sequence space R^∞.
4. The n-dimensional unit cube space $[0,1]^n$, $n = 1, 2, \ldots, \infty$

5. The coin tossing space $\Gamma = \{0,1\}^{\infty}$

6. A compact Hausdorff space with a countable open base. This is homeomorphic and so Borel isomorphic with a closed subset of $[0,1]^{\infty}$.

7. A σ-compact metrizable space, i.e. a metrizable space which is expressed as a countable union of compact subsets. This is standard by Theorem 5.

8. A locally compact Hausdorff space with a countable open base. This is a special case of 7.

9. A complete separable metric space. This is homeomorphic with a G_δ subset of R^{∞}. (See the next section Theorem 5.) Therefore it is a standard Borel space. This includes all the examples mentioned above except 7.

Remark 1. The following theorem is very interesting. Since we shall not use it in this book, we will omit the proof.

Theorem 7. A standard Borel space is Borel isomorphic with one of the following Borel spaces that are obviously not Borel isomorphic with each other.

(a) $\{1,2,3,\ldots,n\}$, $n = 1,2,3,\ldots$

(b) $\{1,2,3,\ldots\}$

(c) $[0,1]$.

1.4 Polish spaces

A topological space is called separable if it has a countable dense subset, and completely metrizable if the topology is determined by a complete metric (= a metric ρ such that every ρ-Cauchy sequence converges to a point in the space). A separable and completely metrizable topological space is called a Polish space.

A complete separable metric space (with the metric topology) is obviously a Polish space. Every topological space homeomorphic with a Polish space is also Polish. Every Polish space obviously has a countable open base.

The real line $R^1 = (-\infty, \infty)$ with the usual topology is Polish. The usual metric $\rho(x,y) = |x-y|$ is complete and the rational numbers form a countable dense set.

The open half real line $R^1_+ = (0, \infty)$ with the usual topology is also Polish, because it is homeormorphic with R^1. The usual metric $\rho(x,y) = |x-y|$ is not complete, because $x_n = 1/n$ is a ρ-Cauchy sequence but does not converge to any point in R^1_+. The metric $\rho_+(x,y) = |\log x - \log y|$ determines the usual topology in R^1_+ and it is complete. The positive rational numbers form a countable dense set in R^1_+. This example shows that a metric space (S, ρ) with the metric topology may be Polish even if the given metric ρ is not complete.

It is easy to see that if $\rho(x,y)$ is a complete metric in S, then $\rho_1(x,y) = \rho(x,y) \wedge 1$ is also a complete metric determining the same topology as the metric ρ. Therefore the topology of a Polish space is determined by a complete metric bounded by 1.

Theorem 1. The product of a countable number of Polish spaces (with the product topology) is Polish.

Proof. We will discuss the countably infinite product case. Let S_n be a Polish space for $n = 1,2,\ldots$ and S the product space $\Pi_n S_n$. Suppose that ρ_n is a complete metric determining the topology in S_n. We can assume that $\rho_n \leq 1$. Then it is easy to check that

$$\rho((x_n),(y_n)) = \sum_n 2^{-n}\rho_n(x_n,y_n)$$

is a complete metric determining the product topology in S.

Let $\{a_{n1}, a_{n2}, \ldots\}$ be a dense subset of S_n. Then the points in S

$$(a_{1k_1}, a_{2k_2}, \ldots, a_{nk_n}, a_{n+1,1}, a_{n+2,1}, \ldots),$$

$$n = 1,2,\ldots, \quad k_i = 1,2,\ldots \quad (i = 1,2,\ldots,n)$$

for a countable dense set, as can be easily checked.

Let us remind the reader of the convention that every subset of a topological space is regarded as a topological space endowed with the relative topology (see 1.2 (a)). A subset of a topological space is called a Polish subset if it is a Polish space with respect to the relative topology.

It is easy to prove the following two propositions.

Proposition 1. A subset of a separable metrizable space is also separable and metrizable.

Proposition 2. A closed subset of a Polish space is Polish.

Proposition 3. An open subset of a Polish space is Polish.

Proof. Let S be a Polish space and ρ_S a complete metric in S. Let G be open in S. G is separable by Proposition 1. We want to define a complete metric ρ_G determining the topology in G. We assume that $S \neq G$, because if $S = G$, then it is enough to set $\rho_G = \rho_S$. Let $f(x)$ be the distance between x and $S-G$, i.e.,

$$f(x) = \inf_{y \in S-G} \rho_S(x,y).$$

Since $|f(x)-f(x')| \leq \rho(x,x')$, f is continuous in S. Since G is open, we have $f(x) > 0$ if and only if $x \in G$. Therefore $g(x) = f(x)^{-1}$ $(x \in G)$ is continuous. Define

$$\rho_G(x,y) = |g(x)-g(y)| + \rho_S(x,y), \qquad x,\ y \in G.$$

This is a metric in G determining the topology in G, as we can easily check.

Let us now prove that ρ_G is complete. Let $\{x_n\}$ be a ρ_G-Cauchy sequence in G. Then $\{x_n\}$ is also a ρ_S-Cauchy sequence by $\rho_S \leq \rho_G$ in G. Therefore we have $x \in S$ such that $x_n \to x$ in S. If $x \in G$, then $x_n \to x$ in G, because the topology in G is the relative topology induced from that in S. Suppose that $x \notin G$. Then $f(x) = 0$. Since f is continuous in S, $\lim_n f(x_n) = f(x) = 0$, so that $\lim_n g(x_n) = \infty$. Thus we have a subsequence $\{y_n\}$ of $\{x_n\}$ such that $g(y_{n+1}) > g(y_n) + 1$. Then $\rho_G(y_n, y_{n+1}) \geq 1$. Since $\{y_n\}$ is a ρ_G-Cauchy sequence as a subsequence of $\{x_n\}$, we must have $\rho_G(y_n, y_{n+1}) \to 0$. This is a contradiction.

Theorem 2. The intersection of a countable number of Polish subsets of a metrizable space is Polish.

Proof. Let S be a metrizable space and $\{T_k\}$ a countable family of Polish subsets of S. We will prove that the intersection $T = \bigcap_k T_k$ is Polish. The product space $P = \prod_k T_k$ with the product topology is Polish. Let $\pi_k: P \to T_k$ is the k-projection that carries $x \in P$ to its kth component. π_k is continuous. The identity map i_k from T_k into S is continuous because of the definition of relative topology. Therefore $\alpha_k = i_k \circ \pi_k: P \to S$ is also continuous. Let ρ be a metric determing the topology in S. Then $(x,y) \to \rho(x,y)$ is a continuous map from S^2 into $[0,\infty)$. Therefore $\xi \to \rho(\alpha_k(\xi),\alpha_k(\xi))$ is a continuous map from P into $[0,\infty)$ and so

$$D = \{\xi \in P: \rho(\alpha_1(\xi),\alpha_k(\xi)) = 0, \ k = 1,2,\ldots\}$$

is a closed subset in P. Since P is Polish by Theorem 1, D is Polish by Proposition 2. But it is easy to see

$$D = \{(x,x,\ldots): x \in T\}.$$

By the definitions of relative topology and product topology we have that for $x_1,x_2,\ldots,x \in T$

$$\begin{aligned} x_n \to x \text{ in } T &\iff x_n \to x \text{ in } S \\ &\iff x_n \to x \text{ in } T_k \text{ for every } k \\ &\iff (x_n,x_n,\ldots) \to (x,x,\ldots) \text{ in } P \\ &\iff (x_n,x_n,\ldots) \to (x,x,\ldots) \text{ in } D. \end{aligned}$$

This implies that $x \to (x,x,\ldots)$ determines a homeomorphic map from T onto D. Since D is Polish, T is also Polish.

<u>Theorem 3</u>. A subset T of a Polish space S is Polish if and only if T is G_δ in S.

<u>Proof</u>. The "if part" follows at once from Proposition 3 and Theorem 2. Let us prove the "only if" part.

Suppose that T is Polish. We will prove that T is G_δ in S. The closure $\overline{T}$ of T is Polish by Proposition 2. Since $\overline{T}$ is G_δ in S as a closed subset of S, it is enough to prove that T is G_δ in $\overline{T}$. Let $\overline{\rho}(\rho)$ be a complete metric in $\overline{T}(T)$. Let $\mathcal{U}_n$ be the family of all $\overline{\rho}$-open subsets U of $\overline{T}$ such that $d_\rho(T \cap U) < 1/n$, where $d_\rho(A)$ is the ρ-diameter, i.e., $\sup\{\rho(x,y)\colon x,y \in A\}$. Set

$$O_n = \bigcup_{U \in \mathcal{U}_n} U.$$

O_n is open in $\overline{T}$. For completion of the proof it is enough to prove that

$$(1) \qquad T = \bigcap_n O_n.$$

Suppose that $x \in T$. Consider the open ball B_n in T with center x and radius $1/2n$. Then $B_n = T \cap U_n$ for some open subset U_n in $\overline{T}$. Since

$$d_\rho(T \cap U_n) = d_\rho(B_n) < 1/n,$$

$U_n \in \mathcal{U}_n$. Therefore $x \in U_n \subset O_n$. As n is arbitrary, $x \in \bigcap_n O_n$.

Suppose that $x \in \bigcap_n O_n$. Then we have $x \in U_n$ for some $U_n \in \mathcal{U}_n$. Let C_n be the open ball in $\overline{T}$ with center x and $\overline{\rho}$-radius $1/n$. Now set

$$V_n = U_1 \cap U_2 \cap \dots \cap U_n \cap C_n.$$

This is open in $\overline{T}$ and contains x. Since T is dense in $\overline{T}$, we have a point $x_n \in T \cap V_n$ for each n. If $n, m > k$, then $x_n, x_m \in T \cap U_k$, and so

$$\rho(x_n, x_m) < d_\rho(T \cap U_k) < 1/k$$

by $U_k \in \mathcal{U}_k$. Since ρ is a complete metric in T, $\{x_n\}$ ρ-converges to a point $y \in T$. Therefore $x_n \to y$ in $\overline{T}$. But $x_n \to x$ in $\overline{T}$ by $x_n \in V_n \subset C_n$. Thus $x = y \in T$. This completes the proof of (1).

Theorem 4. (The Representation Theorem for Polish spaces.) A topological space is Polish if and only if it is homeomorphic with a G_δ subset of the (countably) infinite dimensional unit cube $[0,1]^\infty$.

Proof. As $[0,1]$ is Polish, so is $[0,1]^\infty$. Therefore every G_δ set of $[0,1]^\infty$ is Polish. The "if" part follows at once from this. Let us prove the "only if" part.

Let S be a Polish space and ρ a complete metric bounded by 1 which determines the topology in S. Let $\{a_k\}$ be a sequence dense in S and define a map $S \to f(S)$ $(\subset [0,1]^\infty)$ by

$$f(x) = (\rho(x, a_k),\ k = 1, 2, \ldots).$$

It is easy to check that f defines a homeomorphic map between S and $f(S)$. As S is Polish, so is $f(S)$ and therefore G_δ in $[0,1]^\infty$ by Theorem 3. This completes the proof.

Theorem 5. Every Polish space (with the topological σ-algebra) is a standard Borel space.

Proof. This follows at once from Theorem 4 and 1.3 Example 4.

Examples. The following spaces and their G_δ subsets are Polish.

1. A countable set (with the discrete topology).

2. R^n, $[0,1]^n$, $\{0,1\}^\infty$ $n = 1,2,\ldots,\infty$.

3. A compact Hausdorff space with a countable open base. This is metrizable and separable. Every metric determining the topology is complete.

4. A locally compact Hausdorff space with a countable open base. This is homeomorphic with an open subset of its one point compactification, which is Polish by 3.

1.5 The space C

Let $C = C(I)$, $I = [0,1]$, be the space of all real continuous functions defined on $I = [0,1]$. We introduce a metric ρ_u in C by

(1) $$\rho_u(x,y) = \sup_{t\in I} |x(t) - y(t)| \quad (= \max_{t\in I} |x(t) - y(t)|).$$

It is easy to check that $\rho(x,y)$ satisfies the conditions of a metric. The ρ-topology is often called the uniform convergence topology in view of the fact that $\rho_u(x_n,x) \to 0$ if and only if $x_n(t) \to x(t)$ uniformly in $t \in I$.

Theorem 1. The space $C = C(I)$ with the ρ_u-topology is a Polish space.

Proof. The metric ρ_u itself is complete, because the limit function of a uniformly convergent suquence of continuous functions is also continuous.

To prove that the space C with the ρ_u-topology is separable, we shall find a countable dense subset in C. A function p of the form:

(2) $$\begin{cases} p(t_i) = a_i, \quad i = 0,1,2,\ldots,n \\ p \quad \text{is linear on} \quad [t_{i-1},t_i] \end{cases}$$

$$(0 = t_0 < t_1 < \cdots < t_n = 1, a_0, a_1, \ldots, a_n \in R^1)$$

is called a polygonal function and is denoted by $p_{\{t_i\}\{a_i\}}$. If all t_i and a_i are rational, then $p_{\{t_i\}\{a_i\}}$ is called a rational polygonal function. The rational polygonal functions form a countable set Γ and we can prove that Γ is dense in C using the fact that every $x \in C$ is uniformly continuous.

By 1.4 Theorem 5 and the above theorem we have the following.

Theorem 2. The space C with the topological σ-algebra $\mathcal{B}_u(C)$ relative to the ρ_u topology is a standard Borel space.

For t fixed, the map $e_t: C \to R^1$ defined by $e_t(x) = x(t)$ is called the evaluation map at t. Since

$$|e_t(x) - e_t(y)| = |x(t) - y(t)| \leq \rho(x,y),$$

e_t is continuous and therefore measurable $\mathcal{B}_u(C)/\beta^1$. The σ-algebra on C generated by the sets

$$e_t^{-1}(E),\ t \in I,\ E \in \mathcal{B}^1$$

is called the Kolmogorov σ-algebra on C, $\mathcal{B}_K(C)$ in notation.

Theorem 3. $\mathcal{B}_K(C) = \mathcal{B}_u(C)$.

Proof. Since $e_t^{-1}(E) \in \mathcal{B}_u(C)$ for $E \in \mathcal{B}^1$ and $t \in I$, it is obvious that $\mathcal{B}_K(C) \subset \mathcal{B}_u(C)$. Consider the balls:

$$B(a,r) = \{x \in C: \rho(x,a) \leq r\},\ a \in C,\ r > 0.$$

Since every ρ-open set can be expressed as a countable union of balls, $\mathcal{B}_u(C)$ is the σ-algebra generated by the balls. Let $Q(I)$ be the set of all rationals in I. Then

$$B(a,r) = \bigcap_{t \in Q(I)} \{x \in C: |x(t)-a(t)| \leq r\} \in \mathcal{B}_K(C).$$

This implies that $\mathcal{B}_u(C) \subset \mathcal{B}_K(C)$.

Note. From now on we will write $\mathcal{B}(C)$ for $\mathcal{B}_K(C) = \mathcal{B}_u(C)$.

Theorem 4. The map $\epsilon: I \times C \to R^1$:

$$\epsilon(t,x) = x(t)$$

is continuous in (t,x) and so measurable $\mathcal{B}(I) \otimes \mathcal{B}(C)/\mathcal{B}^1$.

Proof. $I \times C$ is obviously a topological space with the product σ-algebra and the topological σ-algebra $\mathcal{B}(I \times C)$ is identical with the product σ-algebra $\mathcal{B}(I) \otimes \mathcal{B}(C)$ by 1.1 Theorem 1. $\epsilon(t,x)$ is continuous in (t,x) by virtue of

$$|\epsilon(t,x)-\epsilon(t_0,x_0)| \leq |\epsilon(t,x)-\epsilon(t,x_0)| + |\epsilon(t,x_0)-\epsilon(t_0,x_0)|$$

$$\leq \rho_u(x,x_0) + |x_0(t)-x_0(t_0)|.$$

Generalization. The space $C[0,\infty)$ of the real continuous functions defined on $[0,\infty)$ can be dealt with analogously. The metric $\rho(x,y)$ is defined by

$$\rho_u(x,y) = \sum_{n=1}^{\infty} 2^{-n} \sup_{0 \leq t \leq n} \{|x(t)-y(t)| \cap 1\}.$$

The ρ_u-topology is called the topology of uniform convergence on compacts. All the theorems mentioned above hold true for $C[0,\infty)$. Similarly for $C(-\infty,\infty)$.

In the above discussions we have considered real valued functions. Similarly we can discuss the space $C_E(I)$ of all E-valued continuous functions on I where $I = [0,1]$, $[0,\infty)$ or $(-\infty,\infty)$ and E is a complete separable metric space with metric ρ. The only change we have to make is that we should replace $|x(t)-y(t)|$ by $d(x(t),y(t))$. All the theorems mentioned above hold true in this case. Similarly for the case in which E is a Polish space, because E is regarded as a complete separable metric space with an appropriate metric ρ. There may be many such matrics but we can show the ρ_u-topology in $C_E(I)$ is independent of the choice of ρ. We will prove this only in case $I = [0,1]$, as the other cases can be proved similarly with a slight modification. Suppose that this is not true. Then we have

$$\rho_u(x_n, x_0) \to 0 \quad \text{and} \quad \rho_u'(x_n, x_0) \geq c > 0$$

for some ρ and ρ' and some $\{x_n\} \subset E$. Then we have $\rho'(x_n(t_n), x_0(t_n)) \geq c$ for some t_n. We can assume that t_n tends to some t_0. Since x is continuous,

$$\begin{aligned} \rho(x_n(t_n), x_0(t_0)) &\leq \rho(x_n(t_n), x_0(t_n)) + \rho(x_0(t_n), x_0(t_0)) \\ &\leq \rho_u(x_n, x_0) + \rho(x_0(t_n), x_0(t_0)) \\ &\to 0. \end{aligned}$$

Therefore $x_n(t_n) \to x_0(t_0)$ in E and so

$$\rho'(x_n(t_n), x_0(t_0)) \to 0.$$

Since $x_0(t_n) \to x_0(t_0)$, we have

$$\rho'(x_0(t_n), x_0(t_0)) \to 0.$$

This implies $\rho'(x_n(t_n), x_0(t_n)) \to 0$ in contradiction with $\rho'(x_n(t_n), x_0(t_n)) \geq c > 0$.

1.6 The space D

A function $x: I \equiv [0,1] \to R^1$ is called a D function on I if it satisfies the following three conditions:

(D.1) $\lim_{s\downarrow t} x(s) = x(t)$ for every $t \in [0,1]$,

(D.2) $\lim_{s\uparrow t} x(s)$ exists and is finite for every $t \in (0,1]$,

(D.3) $\lim_{s\uparrow 1} x(s) = x(1)$.

The space of all D-functions on I is called the space D on I, D(I) or D in notation.

Every continuous function is a D-function and so C is a subset of D(I).

A step function is called normalized if it is right continuous in [0,1) and left continuous at 1. Every normalized step function is a D function.

Every function of bounded variation normalized in the same sense is a D function.

Theorem 1. Every D function is bounded.

Proof. If a function $x \in D$ is not bounded, we can find a sequence $\{t_n\} \subset [0,1]$ such that

(1) $\lim_n x(t_n) = \infty$.

(2) either $t_1 < t_2 < \cdots$ or $t_1 > t_2 > \cdots$.

Then we have either $x(t-) = \infty$ or $x(t) = \infty$ at $t = \lim_n t_n$ in contradiction with the definition of D functions.

Theorem 2. For $x \in D$, the set

$$A_\epsilon(x) = \{t \in [0,1]: |x(t)-x(t-)| > \epsilon\}$$

is finite and so the set of all discontinuity points of x is countable.

Proof. If $A_\epsilon(x)$ is infinite for some $\epsilon > 0$, we have $\{t_n\} \subset A_\epsilon(x)$ such that

$$t_1 < t_2 < \dots \to t \quad \text{or} \quad t_1 > t_2 > \dots \to t.$$

Then we can find $\{s_n\} \subset [0,1]$ with $|x(t_n)-x(s_n)| > \epsilon$ such that

$$s_1 < t_1 < s_2 < t_2 < \dots \to t \quad \text{or} \quad t_1 > s_1 > t_2 > s_2 > \dots \to t.$$

Then either $\lim_{s\uparrow t} (x(s)$ or $\lim_{s\downarrow t} x(s)$ does not exist in contradiction with the definition of D functions.

Remark. This theorem is an immediate result of Theorem 4 below.

Theorem 3. The limit of a uniformly convergent sequence of D functions is also a D function.

Proof. Obvious by the definition.

Theorem 4. Let x be a D function. For every $\epsilon > 0$ we can find $0 = s_0 < s_1 < \dots < s_n = 1$ such that

$$|x(t)-x(s_i)| \leq \epsilon \quad \text{on} \quad [s_i, s_{i+1}),\ i = 1,2,\dots,n-1.$$

(Then this inequality holds on the closed interval $[s_{n-1}, s_n]$ by the left continuity of x at 1.)

Therefore every D function can be approximated uniformly on $[0,1]$ by normalized step functions.

Proof. Since x is a D function on $I = [0,1]$, for every $u \in [0,1]$ we can find $\delta(u) = \delta(u,\epsilon)$ such that

$$|x(t)-x(u)| < \epsilon \quad \text{for} \quad t \in [u, u+\delta(u)) \cap I$$

and

$$|x(t)-x(u-\delta(u))| < \epsilon \quad \text{for} \quad t \in [u-\delta(u), u) \cap I.$$

Using the covering theorem we have

$$[0,1] \subset \bigcup_{i=1}^{m} (u_i-\delta(u_i),\ u_i + \delta(u_i)) \quad (m < \infty).$$

Let $s_0 < s_1 < \dots < s_n$ be the rearrangement of the points

$$u_i,\ (u_i-\delta(u_i)) \vee 0,\ (u_i+\delta(u_i)) \wedge 1,\ i = 1,2,\dots,m.$$

It is obvious that $s_0 = 0$ and $s_n = 1$. Since $[s_i, s_i+1)$ is included either in some $[u_i-\delta(u_i), u_i)$ or $[u_i, u_i+\delta(u_i))$, we have $|x(t)-x(s_i)| < \epsilon$ on $[s_i, s_{i-1})$.

<u>The Skorohod topology</u>. As in the case of $C(I)$, we can define the <u>supremum metric</u> $\rho(x,y) = \sup_{t\in I}|x(t)-y(t)|$ for $x, y \in D$. This metric determines the <u>uniform convergence topology</u> in D. However, this topology is too strong for many purposes. In fact D is not separable with respect to this topology, because the set of the indicators i_α of $[0,\alpha)$, $0 \leq \alpha \leq 1$ is an uncountable subset of D and we have $\rho(i_\alpha, i_\beta) = 1$ for $\alpha \neq \beta$. A more natural topology with respect to which the space D is Polish was introduced by A. V. Skorohod [1] and is called the <u>Skorohod topology</u> in D.

Let us now define the Skorohod topology. Let Φ be the set of all strongly order preserving maps from $I = [0,1]$ onto itself. Every map $\varphi \in \Phi$ is a homeomorphic map. Φ is a group with respect to the composition of maps. The identity map i plays the unity in the group Φ. It is obvious that $\Phi \subset C \subset D$. For $\varphi \in D$ the transformation $x \to x \circ \varphi$ is 1-1 from D onto itself and preserves the supremum metric $\rho(x,y)$ introduced above, because φ is a 1-1 map from I onto itself. Write $\sigma(\varphi)$ for $\rho(\varphi, i)$ where $\varphi \in \Phi$. It is easy to verify the following properties for $\sigma(\varphi)$:

(σ,1) $0 \leq \sigma(\varphi) < 1$.

(σ,2) $\sigma(\varphi) = 0$ if and only if $\varphi = i$.

(σ,3) $\sigma(\varphi^{-1}) = \sigma(\varphi)$.

(σ,4) $\sigma(\varphi \circ \psi) \leq \sigma(\varphi) + \sigma(\psi)$.

Define the Skorohod metric $\rho_S(x,y)$ in D by

$$\rho_S(x,y) = \inf_{\varphi \in \Phi} [\rho(x, y \circ \varphi) + \sigma(\varphi)].$$

Using the above properties of $\sigma(\varphi)$ we can easily check the conditions of a metric for ρ_S except the property of separation: $\rho_S(x,y) = 0 \Rightarrow x = y$ which will be proved below. Suppose that $\rho_S(x,y) = 0$. Then we have $\varphi_n \in \Phi$ such that $\rho(x \circ \varphi_n, y) \to 0$ and $\sigma(\varphi_n) \to 0$. Therefore $y(t) = \lim_{n\to\infty} x(\varphi_n(t))$ and $\lim_{n\to\infty} \varphi_n(t) = t$. This implies that $y(t) = x(t)$ outside the discontinuity points of x that form a countable set by Theorem 2. Since both x and y are right continuous in $[0,1)$ and left continuous at 1, we have $x(t) = y(t)$ everywhere.

The Skorohod topology is defined to be the ρ_S-topology. Roughly speaking, y is said to be close to x in the Skorohod topology if we can make $y(t)$ be uniformly close to $x(t)$ by little deformation of the t-scale.

Theorem 5. The Skorohod topology is strictly weaker than the uniform convergence topology.

Proof. Since $i \in \Phi$, we have $\rho_S(x,y) \leq \rho(x,y)$. Therefore the ρ_S-topology is weaker than the ρ-topology. Let i_α be the indicator of $[0,\alpha)$ $(0 < \alpha < 1)$. Then $\rho_S(i_\alpha, i_\beta) \leq |\alpha-\beta|$, because if φ is the polygonal function $p_{\{0,\alpha,1\}\{0,\beta,1\}}$ (see 1.5), then $i_\beta \circ \varphi = i_\alpha$ and $\sigma(\varphi) = |\alpha-\beta|$. Therefore $\lim_{\beta\to\alpha} \rho_S(i_\alpha, i_\beta) = 0$, while $\rho(i_\alpha, i_\beta) = 1$ for $\alpha \neq \beta$. This shows that the ρ_S-topology is strictly weaker than the ρ-topology.

Theorem 6. The following functionals on D are continuous with respect to the Skorohod topology:

$$f_1(x) = \sup_t x(t),\ f_2(x) = \inf_t x(t),\ f_3(x) = \sup_t |x(t)|,$$
$$f_4(x) = \sup_t (x(t)-x(t-)),\ f_5(x) = \inf_t (x(t)-x(t-)),$$
$$f_6(x) = \sup_t |(x(t)-x(t-)|.$$

Proof. Suppose that $\rho_S(x_n, x) \to 0$. Then we have $\varphi_n \in \Phi$ such that $y_n(t) \equiv x_n(\varphi_n(t))$ converges to $x(t)$ uniformly in t. Then $f_j(y_n) \to f_j(x)$ for every j. It is obvious by $\varphi_n \in \Phi$ that $f_j(y_n) = f_j(x_n)$. Therefore $f_j(x_n) \to f_j(x)$ for every j.

Theorem 7. The space D with the Skorohod topology is separable.

Proof. Let Γ be the set of all normalized step functions on I whose values and jump points are all rational. Γ is obviously countable. We will prove that Γ is dense in D. Let x be an arbitrary function in D. Then for $\epsilon > 0$ we have a normalized step function y with $\rho(x,y) < \epsilon$ by Theorem 4. We can assume that y takes only rational values. Let $(0 = s_0 <) s_1 < \cdots < s_n (< s_{n+1} = 1)$ be the jump points of y. Let $0 = r_0 < r_1 < \cdots < r_n < r_{n+1} = 1$ be rational numbers such that $|r_i - s_i| < \epsilon$ and write φ for the polygonal function $p_{\{r_i\}\{s_i\}}$. Then $\varphi \in \Phi$ and $\sigma(\varphi) < \epsilon$. Then $z \equiv y \circ \varphi$ belongs to Γ and

$$\rho_S(z,x) \leq \rho(z \circ \varphi^{-1},x) + \sigma(\varphi^{-1}) = \rho(y,x) + \sigma(\varphi) < 2\epsilon.$$

This completes the proof.

The Billingsley metric. The Skorohod metric in D is not complete, as is seen by the following example. Let $x_n = i_{1/n}$, where i_α is the indicator of $[0,\alpha)$. Then $\rho_S(x_m,x_n) \leq |1/m - 1/n|$ as we have seen in the proof of Theorem 5. Therefore $\{x_n\}$ is a ρ_S-Cauchy sequence. But there is no $x \in D$ such that $\rho_S(x_n,x) \to 0$. Suppose that there exists such x. Then $x_n(\varphi_n(t)) \to x(t)$ for some $\varphi_n \in \Phi$ with $|\varphi_n(t)-t| \leq i/n$. Then $x_n(\varphi_n(t)) = 0$ on $(2/n,1]$. Therefore $x(t) = 0$ on $(0,1]$. Since x is right continuous at 0, we have $x(0) = 0$ in contradiction with

$$x(0) = \lim_n x_n(\varphi_n(0)) = \lim_n x_n(0) = 1.$$

The Billingsley metric ρ_B in D that will be defined below is a complete metric determing the Skorohod topology. Let Ψ be the set of all $\varphi \in \Phi$ for which

$$0 < \inf_{t \neq s} \frac{\varphi(t)-\varphi(s)}{t-s} \leq \sup_{t \neq s} \frac{\varphi(t)-\varphi(s)}{t-s} < \infty.$$

Define $\beta(\varphi)$ by

$$\beta(\varphi) = \sup_{t \neq s} \left|\log \frac{\varphi(t)-\varphi(s)}{t-s}\right|, \quad \varphi \in \Phi.$$

Then $\beta(\varphi)$ has the following properties:

(β,1) $\quad 0 \leq \beta(\varphi) \leq \infty.$

(β,2) $\quad \beta(\varphi) = 0$ if and only if $\varphi = i$.

$(\beta,3)$ $\beta(\varphi) < \infty$ if and only if $\varphi \in \Psi$.

$(\beta,4)$ $\beta(\varphi^{-1}) = \beta(\varphi)$.

$(\beta,5)$ $\beta(\varphi_1 \circ \varphi_2) \leq \beta(\varphi_1) + \beta(\varphi_2)$.

$(\beta,6)$ $e^{-\beta(\varphi)}t \leq \varphi(t) \leq e^{\beta(\varphi)}t$.

$(\beta,7)$ $\sigma(\varphi) \leq e^{\beta(\varphi)} - e^{-\beta(\varphi)} \leq 4\beta(\varphi)$ if $\beta(\varphi) \leq 1$.

$(\beta,2)$, $(\beta,3)$, $(\beta,4)$ and $(\beta,5)$ show that Ψ is a subgroup of Φ. The Billingsley metric $\rho_B(x,y)$ is defined by

$$\rho_B(x,y) = \inf_{\psi \in \Psi} [\rho(x, y\circ\psi) + \beta(\psi)].$$

It is easy to check that ρ_B satisfies the conditions of a metric.

Theorem 8. The ρ_B-topology is identical with the Skorohod topology.

Proof. Since both are metric topologies, it is enough to show that "$\rho_B(x_n,x) \to 0$" and "$\rho_S(x_n,x) \to 0$" are equivalent.

If $\rho_B(x_n,x) \to 0$, then we have $\psi_n \in \Psi$ such that

$$\rho(x_n\circ\psi_n, x) \to 0 \quad \text{and} \quad \beta(\psi_n) \to 0.$$

Then $\sigma(\psi_n) \to 0$ by $(\beta,7)$. This implies $\rho_S(x_n,x) \to 0$.

Suppose conversely that $\rho_S(x_n,x) \to 0$. Then we have $\varphi_n \in \Phi$ such that

$$\rho(x_n\circ\varphi_n, x) \to 0 \quad \text{and} \quad \sigma(\varphi_n) \to 0.$$

By Theorem 4 for $\epsilon > 0$ we can find a normalized step function y such that $\rho(y,x) < \epsilon$. Suppose that

$y(t) = a_i$ on $[s_{i-1}, s_i)$, $i = 1,2,\ldots,k$ where $0 = s_0 < s_1 < \cdots < s_k = 1$. Let ψ_n be the polygonal function $p_{\{s_i\},\{\varphi_n(s_i)\}}$. Then $\psi_n \in \Psi$ and

$$\beta(\psi_n) = \max_{1\leq i\leq k} \left|\log \frac{\varphi_n(s_i)-\varphi_n(s_{i-1})}{s_i-s_{i-1}}\right|.$$

Since $|\varphi_n(s_i)-s_i| \leq \sigma(\varphi_n) \to 0$ as $n \to \infty$, we have $\beta(\psi_n) \to 0$. Observe

$$\rho(x_n\circ\psi_n,x) \leq \rho(x_n\circ\psi_n,y) + \rho(y,x),$$

$$\begin{aligned}\rho(x_n\circ\psi_n,y) &\leq \max_i \sup_{t\in I_i} |x_n(\psi_n(t))-a_i|, \quad I_i = [s_{i+1}s_i) \\ &\leq \max_i \sup_{t\in\psi_n(I_i)} |x_n(t)-a_i| \\ &= \max_i \sup_{t\in\varphi_n(I_i)} |x_n(t)-a_i| \quad (\text{by } \psi_n(s_i) = \varphi_n(s_i)) \\ &= \max_i \sup_{t\in I_i} |x_n(\varphi_n(t))-a_i| \\ &= \rho(x_n\circ\varphi_n,y) \\ &\leq \rho(x_n\circ\varphi_n,x) + \rho(x,y),\end{aligned}$$

and so

$$\rho(x_n\circ\psi_n,x) \leq \rho(x_n\circ\varphi_n,x) + 2\rho(x,y).$$

Then we have

$$\limsup_{n\to\infty} \rho(x_n\circ\psi_n,x) \leq 2\rho(x,y) < 2\epsilon.$$

Since ϵ is arbitrary, we have $\lim_n \rho(x_n\circ\psi_n,x) = 0$, which implies

$$\rho_B(x_n,x) \leq \rho(x_n\circ\psi_n,x) + \beta(\psi_n) \to 0 \quad (n\to\infty).$$

This completes the proof.

<u>Theorem 9</u>. The metric ρ_B in D is complete.

<u>Proof.</u> Let $\{x_n\}$ be a ρ_B-Cauchy sequence. We will prove that $\rho(x_n,x) \to 0$ for some $x \in D$. With no loss of generality we can assume that

$$\rho_B(x_n,x_{n+1}) < 2^{-n}, \quad n = 1,2,\ldots .$$

Then we can find $\psi_n \in \Psi$ such that

(1) $\rho(x_n, x_{n+1} \circ \psi_n) < 2^{-n}$ and (2) $\beta(\psi_n) < 2^{-n}$.

Set $\psi_{m,n} = \psi_m \circ \psi_{m-1} \circ \cdots \circ \psi_n$ $(m > n)$. Then we have

$$\begin{aligned}\rho(\psi_{m,n}, \psi_{m+1,n}) &= \rho(i, \psi_{m+1}) \\ &= \sigma(\psi_{m+1}) \\ &< 4\beta(\psi_{m+1}) \leq 2^{-m+1}.\end{aligned}$$

Therefore $\theta_n(t) = \lim_{m\to\infty} \psi_{m,n}(t)$ exists and the convergence is uniform in t for each n. Since $\psi_{n,m} \in \Psi$, θ_n is non-decreasing and continuous. It is obvious that $\theta_n(0) = 0$ and $\theta_n(1) = 1$. But

$$\left|\log \frac{\psi_{m,n}(t) - \psi_{m,n}(s)}{t-s}\right| \leq \beta(\psi_{m,n}) \leq \sum_{k=n}^{m} \beta(\psi_k) < 2^{-n+1}$$

for $t \neq s$. Letting $m \to \infty$, we have

$$\left|\log \frac{\theta_n(t) - \theta_n(s)}{t - s}\right| \leq 2^{-(n+1)} < \infty \quad (t \neq s),$$

which implies that

(2) $\qquad \theta_n \in \Psi$ and $\beta(\theta_n) \leq 2^{-(n+1)} \to 0$.

Observing

$$\begin{aligned}\theta_n(\psi_n^{-1}(t)) &= \lim_m \psi_{m,n}(\psi_n^{-1}(t)) = \lim_n \psi_{m,n+1}(t) \\ &= \theta_{n+1}(t)\end{aligned}$$

i.e. $\quad \theta_n \circ \psi_n^{-1} = \theta_{n+1}$,

we have

$$\begin{aligned}\rho(x_n \circ \theta_n^{-1}, x_{n+1} \circ \theta_{n+1}^{-1}) &= \rho(x_n \circ \theta_n^{-1}, x_{n+1} \circ \psi_n \circ \theta_n^{-1}) \\ &= \rho(x_n, x_{n+1} \circ \psi_n) < 2^{-n}\end{aligned}$$

by (1). Then we can find $x \in D$ such that

(3) $\rho(x_n \circ \theta_n^{-1}, x) \to 0.$

Thus we have $\rho_B(x_n, x) \to 0$ by (2) and (3).

Theorem 10. The space D with the Skorohod topology is Polish.

Proof. D is separable by Theorem 7 and completely metrizable by Theorems 8 and 9.

Let $\mathcal{B}_S(D)$ denote the topological σ-algebra on D relative to the Skorohod topology. By the above Theorem 10 and 1.4 Theorem 5 we have

Theorem 11. The space D with $\mathcal{B}_S(D)$ is a standard Borel space.

The Kolmogorov σ-algebra $\mathcal{B}_K(D)$ on D. Let $e_t\colon D \equiv D(I) \to R^1$ $(I = [0,1])$ be the evaluation map: $e_t(x) = x(t)$. The Kolmogorov σ-algebra $\mathcal{B}_K(D)$ on D is defined to be the σ-algebra generated by

$$e_t^{-1}(E);\ t \in [0,1],\ E \in \mathcal{B}(R^1).$$

We will prove the following important fact.

Theorem 12. $\mathcal{B}_K(D) = \mathcal{B}_S(D)$.

The proof is not as easy as that for the corresponding fact on the space C (1.5 Theorem 3). We shall need some preliminary facts for the proof.

Theorem 13. The map $\epsilon\colon (t,x) \to x(t)$ from $I \times D$ into R^1 is measurable $\mathcal{B}(I) \times \mathcal{B}_K(D)/\mathcal{B}(R^1)$.

Proof. Set $\epsilon_n(t,x) = x(\theta_n(t))$, where

$$\theta_n(t) = \frac{[nt]+1}{n} \wedge 1.$$

Then

$$\epsilon_n(t,x) = x(\tfrac{i}{n}) \quad \text{for} \quad t \in I_{n_i} \equiv [\tfrac{i-1}{n},\tfrac{i}{n})$$
$$= x(1) \quad \text{for} \quad t \in I_{n,n} \equiv [\tfrac{n-1}{n},\tfrac{n}{n}]$$

and therefore ϵ_n is measurable $\mathcal{B}(I) \times \mathcal{B}_K(D)/\mathcal{B}(R^1)$, because

$$\epsilon_n^{-1}(E) = \bigcup_{i=1}^{n} I_{n_i} \times e_{i/n}^{-1}(E).$$

Since $\theta_n(t) \downarrow t$ as $n \to \infty$, we have $\epsilon(t,x) = \lim_n \epsilon_n(t,x)$. Therefore ϵ is also measurable $\mathcal{B}(I) \times \mathcal{B}_K(D)/\mathcal{B}^1$.

Theorem 14. If $\alpha\colon D \to I$ is measurable $\mathcal{B}_K(D)/\mathcal{B}(I)$, then

(a) $x \to x(\alpha(x)) \equiv \epsilon(\alpha(x),x)$ is measurable $\mathcal{B}_K(D)/\mathcal{B}(R^1)$,

and

(b) for $\epsilon > 0$

$$x \to \beta(x) \equiv \inf\{t\colon 1 \geq t > \alpha(x)\colon |x(t)-x(\alpha(x))| > \epsilon\}$$
$$(= 1 \text{ if there is no such } t)$$

is measurable $\mathcal{B}_K(D)/\mathcal{B}(I)$.

Proof.

(a) $x \to (\alpha(x),x)$ is measurable $\mathcal{B}_K(D)/\mathcal{B}(I) \times \mathcal{B}_K(D)$ by 1.1 Theorem 5 and $(t,x) \to \epsilon(t,x)$ is measurable $\mathcal{B}(I) \times \mathcal{B}_K(D)/\mathcal{B}(R^1)$ by the previous theorem. Then their composition $x \to \epsilon(\alpha(x),x)$ is measurable $\mathcal{B}_K(D)/\mathcal{B}(R^1)$.

(b) If $|x(t)-x(\alpha(x))| > \epsilon$, then for every $\delta > 0$ we can find a rational $r \in [t,t+\delta)$ such that $|x(r)-x(\alpha(x))| > \epsilon$. Therefore we can restrict t to rationals in the definition of $\beta(x)$. This implies

$$\{x: \beta(x) < s\} = \bigcup_{\substack{r \text{ rational} \\ r < s}} \{x: r > \alpha(x), |x(r)-x(\alpha(x))| > \epsilon\}.$$

$x \to x(\alpha(x))$ is measurable $\mathcal{B}_K(D)/\mathcal{B}(R^1)$ by (a) and $x \to x(r)$ is measurable $\mathcal{B}_K(D)/\mathcal{B}(R^1)$ be the definition of $\mathcal{B}_K(D)$. Therefore the above set belongs to $\mathcal{B}_K(D)$.

Now we will return to the proof of Theorem 12.

Proof of $\mathcal{B}_K(D) \subset \mathcal{B}_S(D)$. For this purpose it is enough to prove that the evaluation map $e_t(x) \equiv x(t)$ is measurable $\mathcal{B}_S(D)$ for every t. Since $x(\in D)$ is right continuous at $t \in [0,1)$ and left continuous at 1, we have

$$e_t(x) = x(t) = \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon} \int_t^{(t+\epsilon)\wedge 1} x(s)ds \quad (0 \leq t < 1)$$

$$= \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon} \int_{(1-\epsilon)\vee 1}^{1} x(s)ds \quad (t = 1).$$

Therefore it is enough to prove that

$$I_{\alpha\beta}(x) = \int_\alpha^\beta x(s)ds \quad (0 \leq \alpha \leq \beta \leq 1)$$

is continuous in x with respect to the Skorohod topology. Since the Billingsley metric determines the Skorohod topology, it is enough to prove that $\rho_B(x_n,x) \to 0$ implies $I_{\alpha\beta}(x_n) \to I_{\alpha\beta}(x)$. If $\rho_B(x_n,x) \to 0$, then we have $\psi_n \in \Psi$ such that

$$\rho(x_n, x\circ\psi_n) \to 0 \quad \text{and} \quad \beta(\psi_n) \to 0.$$

Then we have

$$(4) \quad |I_{\alpha\beta}(x_n)-I_{\alpha\beta}(x\circ\psi_n)| \leq (\beta-\alpha)\rho(x_n,x\circ\psi_n) \to 0.$$

Observe

$$I_{\alpha\beta}(x\circ\psi_n) = \int_\alpha^\beta x(\psi_n(s))ds$$
$$= \int_{\psi_n(\alpha)}^{\psi_n(\beta)} x(t)d\varphi_n(t) \quad (\varphi_n = \psi_n^{-1})$$
$$= \int_0^1 \lambda_n(t)x(t)d\varphi_n(t),$$

where λ_n is the indicator of the interval $[\psi_n(\alpha),\psi_n(\beta)]$. Since

$$|\psi_n(t)-t| \leq \sigma(\psi_n) \leq e^{\beta(\psi_n)} - e^{-\beta(\psi_n)} \to 0$$

for every t, $\lambda_n(t)$ converges to the indicator $\lambda(t)$ of the interval $[\alpha,\beta]$ except possibly at $t = \alpha,\beta$. Since $\beta(\varphi_n) = \beta(\psi_n) \to 0$, we have

$$0 < e^{-\beta(\varphi_n)} < \frac{\varphi_n(t)-\varphi_n(s)}{t-s} \leq e^{\beta(\varphi_n)} < \infty \quad (t \neq s).$$

This shows that φ_n is absolutely continuous and

$$e^{-\beta(\varphi_n)} \leq \varphi_n'(t) \leq e^{\beta(\varphi_n)} \quad \text{a.e.}$$

Therefore $\varphi_n'(t) \to 1$ a.e. $(n \to \infty)$ by $\beta(\varphi_n) \to 0$. By the bounded convergence theorem we have

$$(5) \qquad I_{\alpha\beta}(x\circ\psi_n) = \int_0^1 \lambda_n(t)x(t)\varphi_n'(t)dt$$
$$\to \int_0^1 \lambda(t)x(t)dt = I_{\alpha\beta}(x).$$

It follows at once from (4) and (5) that $I_{\alpha\beta}(x_n) \to I_{\alpha\beta}(x)$.

Proof of $\mathcal{B}_S(D) \subset \mathcal{B}_K(D)$. Let E be the countable set of all normalized step functions with rational jump points and rational values. If we can construct a sequence of maps $f_n\colon D \to E$ such that

(i) $\rho_S(f_n(x),x) \to 0 \quad (n \to \infty)$ and

(ii) $f_n^{-1}(a) \in \mathcal{B}_K(D)$ for every $a \in E$,

then the ball $B_S(b,r) = \{x \in D: \rho_S(x,b) < r\}$ is expressed as

$$B_S(b,r) = \bigcup_m \bigcup_k \bigcap_{n \geq k} \{x \in D: \rho_S(t_n(x),b) < \frac{m-1}{m} r\}$$

$$= \bigcup_m \bigcup_k \bigcap_{n \geq k} \{x \in D: f_n(x) \in E_m\}$$

where E_m is the countable set of all $a \in E$ with $\rho_S(a,b) < (m-1)r/m$. Since $f_n{}^{-1}(a) \in \mathcal{B}_K(D)$, $f^{-1}(E_m) \in \mathcal{B}_K(D)$. This implies $B_S(b,r) \in \mathcal{B}_K(D)$. Since $\mathcal{B}_S(D)$ is generated by the balls $B_S(b,r)$, $b \in D$, $r > 0$, we have $\mathcal{B}_S(D) \subset \mathcal{B}_K(D)$.

It remains only to construct $\{f_n\}$ with (i) and (ii). Fix n for the moment. We will define $\sigma_k(x)$, $k = 0,1,2,\ldots$ by induction. Set

$$\sigma_0(x) = 0$$

$$\sigma_k(x) = \inf\{t: \sigma_{k-1}(x) < t \leq 1, \ |x(t)-x(\sigma_{k-1}(x))| > \tfrac{1}{n}\}.$$

Define $\sigma_k(x) = 1$ if there is not such t. Since $x \in D$, we have $m = m(x) < \infty$ such that $\sigma_{n-1}(x) < 1$ and $\sigma_m(x) = 1$. Then we have

$$0 = \sigma_0(x) < \sigma_1(x) < \cdots < \sigma_{m-1}(x) < \sigma_m(x) = 1, \ m = m(x).$$

Take the smallest $p = p(x)$ for which

$$p > n \quad \text{and} \quad \frac{1}{p} < \min_{1 \leq i \leq m} (\sigma_i(x)-\sigma_{i-1}(x))$$

and set

$$\tau_0(x) = 0, \ \tau_k(x) = \frac{[p\sigma_k(x)]+1}{p}, \ k = 1,2,\ldots,m-1$$

and

$$\tau_m(x) = 1.$$

Then $0 = \tau_0(x) < \tau_1(x) < \cdots < \tau_m(x) = 1$ and $|\tau_k(x)-\sigma_k(x)| \leq \frac{1}{n}$, $k = 0,1,\ldots,m$. As all $\sigma_k(x)$ are measurable $\mathcal{B}_K(D)$ in x by Theorem 14(a), so are $m(x)$, $p(x)$ and all $\tau_k(x)$.

Notice that $\sigma_k(x)$, $m(x)$, $p(x)$ and $\tau_k(x)$ depend on n. Define $g_n(x) \in D$ and $f_n(x) \in E$ as follows:

$$(g_n(x))(t) = x(\sigma_{k-1}(x)) \quad \text{for} \quad t \in [\sigma_{k-1}(x), \sigma_k(x)),\ k = 1,2,\ldots,m$$

$$= x(\sigma_{m-1}(x)) \quad \text{for} \quad t = 1$$

$$(f_n(x))(t) = \frac{[nx(\sigma_{k-1}(x))]}{n} \quad \text{for} \quad t \in [\tau_{k-1}(x), \tau_k(x)),\ k = 1,2,\ldots m$$

$$= \frac{[nx\sigma_{m-1}(x)]}{n} \quad \text{for} \quad t = 1 .$$

Using the polygonal function $\varphi = p_{\{\sigma_k\}\{\tau_k\}} \in \Phi$, we have

$$\rho_S(f_n(x), g_n(x)) \leq \rho(f_n(x)\circ\varphi) + \sigma(\varphi) < \frac{2}{n}.$$

By the definition of $\sigma_k(x)$ we have

$$\rho_S(g_n(x), x) \leq \rho(g_n(x), x) \leq \frac{1}{n}.$$

Therefore $\rho_S(f_n(x), x) < \frac{3}{n} \to 0 \quad (n \to \infty)$.

For completion of the proof it is enough to show that $f_n^{-1}(a) \in \mathcal{B}_K(D)$ for every $a \in E$. Suppose that $a(t) = \alpha_1, \alpha_2, \ldots$ or α_g according as $t \in [0, t_1), [t_1, t_2), \ldots,$ or $[t_{q-1}, 1]$. If all of $n\alpha_k$, $k = 1,2,\ldots,q-1$ are integers, $f_n^{-1}(a)$ is the set of all $x \in D$ satisfying the conditions

$$m(x) = q,\ \tau_k(x) = t_k \quad (k = 1,2,\ldots,q-1),$$

and

$$\alpha_k \leq x(\sigma_{k-1}(x)) < \alpha_k + \frac{1}{n} \quad (k = 1,2,\ldots,q).$$

Since $m(x)$ and $\tau_k(x)$ are measurable $\mathcal{B}_K(D)$ in x and $x(\sigma_{k-1}(x))$ are all measurable $\mathcal{B}_K(D)$ in x, by Theorem 14(b) we have $f_n^{-1}(a) \in \mathcal{B}_K(D)$. If at least one of $n\alpha_k$, $k = 1,2,\ldots,q-1$ is not an integer, $f_n^{-1}(a) = \emptyset \in \mathcal{B}_K(D)$.

The space C is a subspace of the space D. The space C is endowed with the uniform convergence topology τ_u and the space D is endowed with the Skorohod topology τ_S.

Theorem 15. C is closed in D and τ_u is identical with the relative topology in C induced from τ_S in D.

Proof. Using the functional $f_6(x)$ in Theorem 6, we have $C = f_6^{-1}(0)$. Since f_6 is continuous in D, C is closed in D. For the proof of the second part it is enough to show that $\rho(x_n,x) \to 0$ and $\rho_S(x_n,x) \to 0$ are equivalent for $x_1,x_2,\ldots,x \in C$. Since $i \in \Phi$, $\rho_S(x_n,x) \leq \rho(x_n,x)$. Therefore $\rho(x_n,x) \to 0$ implies $\rho_S(x_n,x) \to 0$. Suppose that $\rho_S(x_n,x) \to 0$. Then we have $\varphi_n \in \Phi$ such that $\rho(x_n,x\circ\varphi_n) \to 0$ and $\sigma(\varphi_n) \to 0$. Since $x(t)$ is uniformly continuous on $[0,1]$ and since $\varphi_n(t) \to t$ uniformly on $[0,1]$ by $\sigma(\varphi_n) \to 0$, $\rho(x\circ\varphi_n,x) \to 0$. Therefore $\rho(x_n,x) \to 0$. This completes the proof.

Generalization. As in the case of the space C, we can consider the space $D[0,\infty)$ of all right continuous functions with finite left limit at every $t \in (0,\infty)$. It is a Polish space with respect to the topology defined by the metric

$$\rho_S(x,y) = \sum_{n=1}^{\infty} 2^{-n}(\rho_S(x_n,y_n) \cap 1)$$

where $x_n(y_n)$ is the restriction of $x(y)$ to $[0,n]$ modified at n by $x_n(n) = x(n-)$ $(y_n(n) = y(n-))$ and the metric $\rho_S(x_n, y_n)$ is defined exactly in the same way as in $D[0,1]$. Similarly for $D(-\infty, \infty)$.

The extension to the case $D_E(I)$ in which $I = [0,1]$, $[0,\infty)$ or $(-\infty,\infty)$ and E is a Polish space can be discussed exactly in the same way as in the case of C.

1.7 The space M of canonical measurable functions

Let I be the closed unit interval [0,1] and $\widetilde{M} = \widetilde{M}(I)$ the space of all real measurable functions defined a.e. on I. Whenever we consider the space $\widetilde{M}$ or a subspace of $\widetilde{M}$ such as $L^p(I)$, we identify two equivalent (= equal a.e.) functions in $\widetilde{M}$. Therefore $\widetilde{M}$ is the space of equivalence classes rather than the space of functions.

The spaces C(I) and D(I) are subspaces of $\widetilde{M}$. Since C(I) and D(I) are spaces of functions and $\widetilde{M}$ is a space of equivalence classes, the above statement needs some interpretation. Let $\widetilde{C}(I)$ be the set of all equivalence classes in $\widetilde{M}$ represented by functions in C(I). By continuity of x in C(I) two different $x,y \in C(I)$ determine different equivalence classes in $\widetilde{M}$. Therefore the natural correspondence $C(I) \to \widetilde{C}(I)$ is 1-1 and we can identify $\widetilde{C}(I)$ with C(I). It is in this sense that C(I) is regarded as a subspace of $\widetilde{M}$. Similarly for D(I).

To do without such an interpretation, we will pick up the most regular (in some sense) function, called a <u>canonical measurable function</u>, from each equivalence class and consider the space M = M(I) of all canonical measurable functions instead of $\widetilde{M}$. Then we have

$$C(I) \subset D(I) \subset M(I)$$

in the naive set-theoretical sense.

Before defining canonical measurable functions we will review the classical notion of <u>approximate limit</u>.

Definition 1. Let x be a function in $\widetilde{M}$ and t a point in $(0,1)$. If for every neighborhood U of a we have

$$(1) \qquad \lim_{\epsilon_1,\epsilon_2 \downarrow 0} \frac{m(x^{-1}(U) \cap [t-\epsilon_1, t+\epsilon_2])}{m[t-\epsilon_1, t+\epsilon_2]} = 1$$

then we call a the approximate limit of $x(s)$ as $s \to t$, $\text{a-lim}_{s\to t}x(s)$ in notation.

Note that such a is unique if it exists. Suppose (1) is true for $a = a_1$ and $a_2 (a_1 \neq a_2)$. Take neighborhoods $U_1 = U_1(a_1)$ and $U_2 = U_2(a_2)$ such that $U_1 \cap U_2 = \emptyset$. Then $x^{-1}(U_1) \cap x^{-1}(U_2) = \emptyset$ and therefore (1) holds for $U = U_1$ and U_2 and so we have

$$\lim_{\epsilon_1,\epsilon_2 \downarrow 0} \frac{m((x^{-1}(U_1) \cup x^{-1}(U_2)) \cap [t-\epsilon_1, t+\epsilon_2])}{m[t-\epsilon_1, t+\epsilon_2]} = 2,$$

but the ratio must be always ≤ 1. This is a contradiction.

For $t \in [0,1)$ (or $(0,1]$) we will define the approximate right limit $\text{a-lim}_{s\downarrow t}\, x(s)$ (or the approximate left limit $\text{a-lim}_{s\uparrow t}x(s)$) by restricting ϵ_1 (or ϵ_2) to be identically 0 in (1).

It is easy to see that $\text{a-lim}_{s\to t}x(s)$ exists if and only if both $\text{a-lim}_{s\downarrow t}x(s)$ and $\text{a-lim}_{s\uparrow t}x(s)$ exists and take the same value.

Corollary. If $x = y$ a.e., then

$$\text{a-lim}_{s\to t}x(t) = \text{a-lim}_{s\to t}y(t),$$

i.e. if one of the limits exists, then the other one exists and both limits are the same. Similarly for the approximate right (or left) limit at $t \in [0,1)$ (or $t \in (0,1]$).

Proof. Obvious by the definition.

Theorem 1. All the three approximate limits exist a.e. on I and are equal to the original value $x(t)$ a.e. on I.

Proof. It is enough to find a set I' with $m(I - I') = 0$ such that $\text{a-lim}_{s\to t} x(t)$ exists and equals $x(t)$ for every $t \in I'$. Let $\{a_k\}$ be a countable dense subset in I. Let U_{nk} be the $1/n$-neighborhood of a_k, $I_{nk} = x^{-1}(U_{nk})$ and I'_{nk} the set of all $t \in I_{nk}$ for which

$$\lim_{\epsilon_1, \epsilon_2 \downarrow 0} \frac{m(I_{nk} \cap [t-\epsilon_1, t+\epsilon_2])}{m[t-\epsilon_1, t+\epsilon_2]} = 1.$$

By the density theorem we have

$$m(I_{nk} - I'_{nk}) = 0.$$

Since $\bigcup_k I_{nk}$ is the same as the set $\mathcal{D}(x)$ of all t for which $x(t)$ is defined, we have

$$m(I - \bigcup_k I_{nk}) = 0.$$

Therefore $m(I - \bigcup_k I'_{nk}) = 0$ and so

$$m(I - \bigcap_n \bigcup_k I'_{nk}) = 0.$$

Setting $I' = \bigcap_n \bigcup_k I'_{nk}$, we have $m(I - I') = 0$.

We will prove that $\text{a-lim}_{s\to t} x(t)$ exists and equals $x(t)$ for every $t \in I'$. For $t \in I'$, we can find $I'_{n,k(n)}$, $n = 1,2,\ldots$ such that

$$t \in I'_{n,k(n)}, \quad n = 1,2,\ldots\,.$$

Then it holds that

$$(2) \qquad \lim_{\epsilon_1,\epsilon_2\downarrow 0} \frac{m(x^{-1}(U_{nk(n)}) \cap [t-\epsilon_1, t+\epsilon_2])}{m[t-\epsilon_1, t+\epsilon_2]} = 1.$$

Since $t \in I'_{n,k(n)} \subset I_{n,k(n)}$, we have

$$x(t) \in U_{n,k(n)} \quad \text{and so} \quad |x(t) - a_{k(n)}| \leq 1/n.$$

Therefore $U_{n,k(n)}$ is included in the $2/n$-neighborhood of $x(t)$. For every neighborhood U of $x(t)$ we can find $U_{n,k(n)} \subset U$ by taking n big enough. Then we have

$$\liminf_{\epsilon_1,\epsilon_2\downarrow 0} \frac{m(x^{-1}(U) \cap [t-\epsilon_1, t+\epsilon_2])}{m[t-\epsilon_1, t+\epsilon_2]}$$

$$\geq \lim_{\epsilon_1,\epsilon_2\downarrow 0} \frac{m(x^{-1}(U_{n,k(n)}) \cap [t-\epsilon_1, t+\epsilon_2])}{m[t-\epsilon_1, t+\epsilon_2]} = 1,$$

which implies that $\text{a-lin}_{s\downarrow t} x(t)$ exists and equals $x(t)$.

<u>Theorem 2</u>. $\text{a-}\lim_{s\to t} x(s) = a$

if and only if

$$(3) \qquad \lim_{\epsilon_1,\epsilon_2\downarrow 0} \frac{1}{\epsilon_1+\epsilon_2} \int_{t-\epsilon_1}^{t+\epsilon_2} \rho(x(s),a)\,ds = 0$$

where $\rho(a,b) = |a-b| \wedge 1$. Similar facts hold for the other approximate limits.

<u>Proof</u>. Let U_η be the η-neighborhood of a $(0 < \eta < 1)$. Then

$$\frac{1}{\epsilon_1+\epsilon_2}\int_{t-\epsilon_1}^{t+\epsilon_2}\rho(x(s),a)ds \geq \eta\left(1-\frac{m(x^{-1}(U_\eta)\cap[t-\epsilon_1,t+\epsilon_2])}{m[t-\epsilon_1,t+\epsilon_2]}\right)$$

$$\leq \eta+\left(1-\frac{m(x^{-1}(U_\eta)\cap[t-\epsilon_1,t+\epsilon_2])}{m[t-\epsilon_1,t+\epsilon_2]}\right)$$

from which our theorem follows at once.

Definition 2. For $x \in \widetilde{M}$ we define the canonical modification $x_c(t)$ as follows:

(4) $$x_c(t) = \begin{cases} \text{a-lim}_{s\downarrow t}x(s) & (0 \leq t < 1) \\ \text{a-lim}_{s\uparrow 1}x(s) & (t = 1) \end{cases}$$

if this limit exists and $x_c(t)$ is undefined elsewhere.

By Theorem 1 $x_c(t)$ is defined a.e. on I and equals $x(t)$ a.e. on I. Therefore x_c belongs to the same equivalence class as x. If x and y belongs to the same equivalence class, i.e. $x = y$ a.e., then $x_c = y_c$, i.e. x_c and y_c are defined on the same set and take the same values on the set. Since $x_c = x$ a.e., we have $(x_c)_c = x_c$. In view of this fact we define the canonical measurable functions as follows.

Definition 3. If $x = x_c$, then x is called a canonical measurable function.

Then we have one and only one canonical measurable function in each equivalence class.

If $\lim_{s\downarrow t}x(s)$ exists, then $\text{a-lim}_{s\downarrow t}x(s)$ exists and these limits are equal. Similarly for $\lim_{s\uparrow t}x(s)$ and $\text{a-lim}_{s\uparrow t}x(s)$.

Therefore all functions in $D = D(I)$ and so all functions in $C = C(I)$ are canonical measurable functions. Thus we have the following.

Theorem 3. $C \subset D \subset M$.

Now we will topologize $M = M(I)$ by the metric

$$\rho_m(x,y) = \int_I \rho(x(t),y(t))dt, \quad \rho(a,b) = |a-b| \wedge 1.$$

It is obvious that $\rho_m(x,x) = 0$ and $\rho_m(x,y) = \rho_m(y,x)$. Since $\rho(a,b) + \rho(b,c) \geq \rho(a,c)$, we have

$$\rho_m(x,y) + \rho_m(y,z) \geq \rho_m(x,z).$$

If $\rho_m(x,y) = 0$, then $x(t) = y(t)$ a.e. on I. Therefore $x = x_c = y_c = y$ by virtue of $x,y \in M$. This proves that ρ_m satisfies all conditions of a metric.

Since for $0 < \eta < 1$ we have

$$\eta\, m\{t: |x(t)-y(t)| > \eta\} \leq \rho_m(x,y) \leq \eta + m\{t: |x(t)-y(t)| > \eta\}$$

the ρ_m-topology is identical with the topology of conveyence in measure.

The topological σ-algebra in M (with respect to the ρ_m-topology) is denoted by $\mathcal{B}_m(M)$.

Theorem 4. The space M with the ρ_m-topology is a Polish space and therefore a standard Borel space with $\mathcal{B}_m(M)$.

Proof. First we will prove that M is separable. Let A be a countable dense subset in R^1 and Λ the set of all functions $\xi: I \to R^1$ of the following form:

$$\xi(t) = a_i \quad \text{on} \quad [r_{i-1}, r_i) \qquad (1 \leq i \leq n-1)$$
$$= a_n \quad \text{on} \quad [r_{n-1}, r_n],$$

where $a_i \in A$, r_i is rational for every i and

$$0 = r_0 < r_1 < \dots < r_n = 1.$$

Λ is obviously a countable subset of M. We will prove that Λ is dense in M. It is enough to prove that for $x \in M$ and $\epsilon > 0$ we can find $\xi \in \Lambda$ such that $\rho_m(x, \xi) < 5\epsilon$. Write I_k for the t-set:

$$\rho(x(t), a_i) \geq \epsilon \quad \text{for} \quad i < k$$

and

$$\rho(x(t), a_k) < \epsilon.$$

Then I_k, $k = 1, 2, \dots$ are disjoint, $\rho(x(t), a_k) < \epsilon$ for $t \in I_k$ and $\bigcup_k I_k$ is the set where x is defined, so that

$$m(I - \bigcup_k I_k) = 0.$$

Take $K = K(\epsilon)$ such that $m(I - \bigcup_{k=1}^{K} I_k) < \epsilon$. A set expressed as a finite disjoint union of rational intervals of the form $[r, r')$ $(0 \leq r < r' < 1)$ or $[r, 1]$ $(0 \leq r < 1)$ will be called an elementary set for the moment. We denote by $\mathcal{E}$ the class of all elementary sets. Since I_k is measurable, we can find $J_k \in \mathcal{E}$ such that

$$m(I_k \,\theta\, J_k) < \epsilon/K^2, \quad k = 1,2,\ldots,K, \quad A\theta B = (A-B)\cup(B-A)$$

Since $I_k \cap I_j = \emptyset$ for $k \neq j$, we have

$$m(J_k \,\theta\, J_j) = m((J_k \cap J_j) \,\theta\, (I_k \cap I_j))$$

$$\leq m(J_k \,\theta\, I_k) + m(J_j \,\theta\, I_j) < 2\epsilon/K^2.$$

Set

$$J_k' = J_k - \bigcup_{j=1}^{k-1} (J_k \cap J_j) \in \mathcal{E}.$$

Then

$$m(J_k \,\theta\, J_k') \leq \sum_{j=1}^{k-1} m(J_k \cap J_j) < 2\epsilon/K$$

and so

$$m(I_k \,\theta\, J_k') < m(I_k \,\theta\, J_k) + m(J_k \,\theta\, J_k') < 3\epsilon/K.$$

Define ξ by

$$\xi(t) = a_k \quad \text{on} \quad J_k', \; k = 1,2,\ldots,K$$

$$= a_0 \text{ (any fixed point in } R^1\text{, say 0) elsewhere.}$$

Since $J_k' \in \mathcal{E}$, it is easy to check that $\xi \in \Lambda$.

Observing

$$m(I - \bigcup_k I_k \cap J_k')$$

$$\leq m(I - \bigcup_k I_k) + \sum_k m(I_k \,\theta\, J_k')$$

$$< \epsilon + \frac{3\epsilon}{K} \cdot K = 4\epsilon,$$

we obtain

$$\rho_m(x,\xi) = \sum_{k=1}^{K} \int_{I_k \cap J'_k} \rho(x(t),a_k)dt + \int_{I-\bigcup_k I_k \cap J'_k} \rho(x(t),\xi(t))dt$$

$$< \epsilon + m(I - \bigcup_k I_k \cap J'_k) < 5\epsilon.$$

This completes the proof of separability of M.

Second we will prove that ρ_m is a complete metric. For this purpose it is enough to prove that if $\rho_m(x_p,x_{p+1}) \leq 2^{-p}$, then we have $x \in M$ such that $\rho_m(x_p,x) \to 0$ as $p \to \infty$. Observing

$$\int_I \sum_p \rho(x_p(t),x_{p+1}(t))dt = \sum_p \rho_m(x_p,x_{p+1}) \leq 1,$$

we have

$$\sum_p \rho(x_p(t),x_{p+1}(t)) < \infty \quad \text{a.e. on } I,$$

so that

$$\rho(x_p(t),x_q(t)) \leq \sum_{k \geq p \wedge q} \rho(x_k(t),x_{k+1}(t))$$

$$\to 0 \quad \text{a.e. on } I$$

as $p,q \to \infty$, and so

$$\lim_{p,q\to\infty} |x_p(t) - x_q(t)| = 0 \quad \text{a.e. on } I.$$

Then we can find $x \in \widetilde{M}$ such that

$$\lim_{p\to\infty} |x_p(t) - x(t)| = 0 \quad \text{a.e. on } I.$$

By replacing x by its canonical modification, we have $x \in M$ with the above property. By the bounded convergence theorem we have

$$\lim_{p\to\infty} \rho_m(x_p,x) = \lim_{p\to\infty} \int_I |x_p(t) - x(t)| \wedge 1 \, dt$$

$$= \int_I \lim_{p\to\infty} |x_p(t) - x(t)| \wedge 1 \, dt = 0.$$

We have already seen $C \subset D \subset M$ in Theorem 3. Now we will prove the following.

Theorem 5. $C, D \in \mathcal{B}_m(M)$, $\mathcal{B}_u(C) = C \cap \mathcal{B}_m(M)$ and $\mathcal{B}_s(D) = D \cap \mathcal{B}_m(M)$. Therefore

$$\mathcal{B}_u(C) = \{\Lambda : \Lambda \subset C, \ \Lambda \in \mathcal{B}_m(M)\}$$

$$\mathcal{B}_s(D) = \{\Lambda : \Lambda \subset D, \ \Lambda \in \mathcal{B}_m(M)\}.$$

Proof. Since we have proved $C \in \mathcal{B}_s(D)$ and $\mathcal{B}_u(C) = C \cap \mathcal{B}_s(D)$ in 1.6 Theorem 15, it is enough to prove that $D \in \mathcal{B}_m(M)$ and $\mathcal{B}_s(D) = D \cap \mathcal{B}_m(M)$. We will write $\mathcal{D}(x)$ for the set where $x (\in M)$ is defined. It is needless to say that $m(I - \mathcal{D}(x)) = 0$.

First we will make two small remarks.

(1) If $t \in \mathcal{D}(x)$, then we can find an arbitrarily small rational interval J containing t such that

$$\frac{1}{m(J)} \int_J \rho(x(s), x(t)) ds$$

is arbitrarily small. This follows from the definition of canonical measurable functions, the second part of Theorem 2 and the fact that the above quantity depends continuously on the end points of J.

(2) If $t, s \in \mathcal{D}(x)$, then we can find arbitrarily small rational intervals $J \ni t$ and $K \ni s$ such that

$$\frac{1}{m(J)m(K)} \int_J \int_K \rho(x(u),x(v))du\ dv$$

is arbitrarily close to $\rho(x(s),x(t))$. To prove this, observe

$$\left|\frac{1}{m(J)m(K)} \int_J \int_K \rho(x(u),x(v))du\ dv - \rho(x(s),x(t))\right|$$

$$\leq \frac{1}{m(J)m(K)} \int_J \int_K |\rho(x(u),x(v)) - \rho(x(s),x(t))|du\ dv$$

$$\leq \frac{1}{m(J)m(K)} \int_J \int_K [\rho(x(u),x(s)) + \rho(x(v),x(t))]du\ dv$$

$$= \frac{1}{m(J)} \int_J \rho(x(u),x(s))du + \frac{1}{m(K)} \int_K \rho(x(u),x(t))dv$$

and use (1).

Let us consider the supremum $N_p(x)$ if the number n for which

(3) there exist $2n$ points $s_1,t_1,s_2,t_2,\ldots,s_n,t_n \in \mathcal{D}(x)$

such that

$$0 \leq s_1 < t_1 < \ldots < s_n < t_n < 1$$

and

$$\rho(x(s_i),x(t_i)) > 1/p,\ i = 1,2,\ldots n.$$

The condition (3) implies the following condition by the remark (2):

(4) there exist $2n$ rational intervals $J_1,K_1,\ldots,J_n,K_n \subset I$ such that

$$J_1 < K_1 < J_2 < K_2 < \dots < J_n < K_n$$

and

$$\frac{1}{m(J_i)m(K_i)} \int_{J_i} \int_{K_i} \rho(x(s),x(v))du\, dv > 1/p,$$

$$i = 1,2,\dots,n,$$

where $J < J'$ means that the right end of J is smaller than the left end of J'. Conversely if (4) holds, we can find $s_i \in \mathcal{D}(x) \cap J_i$, and $t_i \in \mathcal{D}(x) \cap K_i, i=1,2,\dots$ for which we have $\rho(x(s_i),x(t_i)) > 1/p$. Therefore (3) and (4) are equivalent.

Now we will prove that $N_p(x)$ is measurable $\mathcal{B}_m(M)$ in x, namely that the set $\{x:N_p(x) \geq n\}$ belongs to $\mathcal{B}_m(M)$. This set is characterized by the condition (3) i.e. (4). The system of rational intervals $J_1,K_1,\dots,J_n,K_n$ with $J_1 < K_1 < J_2 < K_2 < \dots < J_n < K_n$ is obviously countable. Therefore for the proof that $\{x:N_p(x) \geq n\} \in \mathcal{B}_m(M)$ it is enough to prove that

$$\mu(x) = \frac{1}{m(J)m(K)} \int_J \int_K \rho(x(u),x(v))du\, dv$$

is measurable $\mathcal{B}_m(M)$ in x. As a matter of fact, $\mu(x)$ is continuous in x, because we can easily see that

$$|\mu(x) - \mu(y)| \leq \frac{1}{m(J)} \int_J \rho(x(u),y(u))du + \frac{1}{m(K)} \int_J \rho(x(u),y(v))dv$$

$$\leq \frac{1}{m(J)} \rho_m(x,y) + \frac{1}{m(K)} \rho_m(x,y).$$

Thus the proof of measurability $\mathcal{B}_m(M)$ of $N_p(x)$ is completed.

For the proof that $D \in \mathcal{B}_m(M)$ it is enough to verify the following identity.

(5) $$D = \bigcap_p D_p, \quad D_p = \{x : N_p(x) < \infty\}.$$

Suppose that $x \in D$. Then we can find a normalized step function x_p such that

$$\sup_{t \in I} \rho(x(t), x_p(t)) < \frac{1}{2p} \quad .$$

This implies that $N_p(x) \leq$ the number of jumps of $x_p < \infty$, so that $x \in D_p$ for every p.

Supposing conversely that $x \in \bigcap_p D_p$, we will prove $x \in D$. Take an arbitrary $t \in [0,1)$. Then

(6) $$b = \lim_{\substack{s \downarrow t \\ s \in \mathcal{D}(x)}} x(s) \quad \text{exists and is finite,}$$

because if otherwise, we can find two sequences $\{s_n\}$ and $\{t_n\}$ in $\mathcal{D}(x)$ such that $s_1 > t_1 > s_2 > t_2 > \ldots \to t$ and $\rho(x(s_n), x(t_n)) > 1/p$ for some p independent of n and therefore we have $x \notin D_p$ in contradiction with our assumption. K follows from (6) by the bounded convergence theorem that

$$\frac{1}{\epsilon} \int_t^{t+\epsilon} \rho(x(s), b) ds = \int_0^1 \rho(x(t+\epsilon s, b) ds \to 0;$$

recall here that $m(I - \mathcal{D}(x)) = 0$. This implies that $t \in \mathcal{D}(x)$ and

(7) $$x(t) = b = \lim_{\substack{s \downarrow t \\ s \in \mathcal{D}(x)}} x(s) \ .$$

Therefore $[0,1) \subset \mathcal{D}(x)$. Then (7) can be written as

(8) $$x(t) = \lim_{s \downarrow t} x(s).$$

Similarly we can show that (i) $\lim_{s \uparrow t} x(s)$ exists and is finite for $s \in [0,1)$ and (ii) $1 \in \mathcal{D}(x)$ and $\lim_{s \uparrow 1} x(s) = x(1)$. This completes the proof of (5).

For completion of the proof of our theorem it is enough to show that $\mathcal{B}_S(D) = D \cap \mathcal{B}_m(M)$. The right hand side is the topological σ-algebra in D with respect to the ρ_m-topology <u>in</u> D and so we will write it as $\mathcal{B}_m(D)$.

If $\rho_S(x_n, x) \to 0$, then we have $\varphi_n \in \Phi$ such that $x_n(t) - (x \circ \varphi_n)(t) \to 0$ and $\varphi_n(t) \to t$ uniformly in t. Then $(x \circ \varphi_n)(t) \to x(t)$ at all continuity points of x. Since the discontinuity points of x form a countable set, $(x \circ \varphi_n)(t) \to x(t)$ a.e. on I. Therefore

$$\rho_m(x_n, x) \leq \rho_m(x_n, x \circ \varphi_n) + \rho_m(x \circ \varphi_n, x)$$

$$\to 0 \quad (n \to \infty).$$

Therefore the ρ_S-topology in D is stronger than the ρ_m-topology in D. This implies that $\mathcal{B}_S(D) \supset \mathcal{B}_m(D)$.

Since $\mathcal{B}_K(D) = \mathcal{B}_S(D)$ by <u>1.6</u> Theorem 12, for the proof of $\mathcal{B}_S(D) \subset \mathcal{B}_m(D)$ it is enough to show that $\mathcal{B}_K(D) \subset \mathcal{B}_m(D)$, i.e. $e_t(x) = x(t)$ is measurable $\mathcal{B}_m(D)$ in $x \in D$ for every $t \in I$. Since

$$\left|\int_J \rho(x(s),a)ds - \int_J \rho(y(s),a)ds\right| \leq \int_J \rho(x(s),y(s))ds \leq \rho_m(x,y),$$

$\int_J \rho(x(s),a)ds$ is ρ_m-continuous and so measurable $\mathcal{B}_m(D)$ in $x \in D$ for every $a \in R^1$. Since

$$\rho(x(t),a) = \lim_{n\to\infty} n \int_t^{t+\frac{1}{n}} \rho(x(s),a)ds, \quad t \in [0,1)$$

$$= \lim_{n\to\infty} n \int_{1-\frac{1}{n}}^{1} \rho(x(s),a)ds, \quad t = 1$$

for $x \in D$ and $a \in R^1$, $\rho(x(t),a)$ is measurable $\mathcal{B}_m(D)$ in x for every $a \in R^1$. Let $\{a_k\}$ be a dense sequence in R^1 and define $e_t^n(x)$ to be the first a_k such that $\rho(x(t),a_k) < 1/n$. Then $e_t^n(x)$ is clearly measurable $\mathcal{B}_m(D)$. Since $\rho(x(t),e_t^n(x)) < 1/n \to 0$, $e_t(x) = \lim_{n\to\infty} e_t^n(x)$ for every x. This proves that $e_t(x)$ is also measurable $\mathcal{B}_m(D)$ in $x \in D$. Thus we have proved $\mathcal{B}_S(D) = \mathcal{B}_K(D) = \mathcal{B}_m(D)$, which completes the proof of our theorem.

Consider the map $e_t: x \to x(t)$ and $\epsilon: (t,x) \to x(t)$. The domain $\mathcal{D}(e_t)$ of definition of e_t is the set of all $x \in M$ for which $x(t)$ is defined, while the domain of definition $\mathcal{D}(\epsilon)$ of ϵ is the set of all $(t,x) \in I \times M$ for which $x(t)$ is defined. The product space $I \times M$ is a topological space with the product topology. The topological σ-algebra $\mathcal{B}(I \times M)$ is identical with the product σ-algebra $\mathcal{B}(I) \otimes \mathcal{B}_m(M)$ by 1.1 Theorem 1.

Theorem 6. Both e_t and ϵ are Borel measurable. More precisely,
(a) for every t, $\mathcal{D}(e_t) \in \mathcal{B}_m(M)$ and $e_t : \mathcal{D}(e_t) \to R^1$ is measurable $\mathcal{D}(e_t) \cap \mathcal{B}_m(M)/\mathcal{B}(R^1)$.
(b) $\mathcal{D}(\epsilon) \in \mathcal{B}(I \times M)$ and $\epsilon : \mathcal{D}(\epsilon) \to R^1$ is measurable $\mathcal{D}(\epsilon) \cap \mathcal{B}(I \times M)/\mathcal{B}(R^1)$.

Proof. (a) follows from (b) by 1.1 Theorem 3. To prove (b), let us observe the following functions:

$$f_\delta(t,x,a) = \frac{1}{\epsilon} \int_t^{t+\delta} \rho(x(s),a)ds \qquad 0 \leq t \leq 1-\delta$$

$$= \frac{1}{\epsilon} \int_{1-\epsilon}^{1} \rho(x(s),a)ds \qquad 1-\delta < t \leq 1,$$

$$f(t,x,a) = \limsup_{\delta \downarrow 0} f_\delta(t,x,a) = \lim_{n\to\infty} \sup_{0<\delta<\frac{1}{n}} f_\delta(t,x,a)$$

and

$$f(t,x) = \inf_{a \in R^1} f(t,x,a).$$

Since $\rho \leq 1$ and since

$$|\rho(x(s),a) - \rho(y(s),b)| \leq \rho(x(s),y(s)) + \rho(a,b),$$

we can easily verify

(1) $$|f_\delta(t,x,a) - f_\delta(u,y,b)|$$
$$\leq 2|t-u| + \frac{1}{\delta}\rho_m(x,y) + \rho(a,b).$$

As special cases of this inequality we have

(1.a) $\quad |f_\delta(t,x,a) - f_\delta(u,y,a)| \leq 2|t-u| + \frac{1}{\delta}\rho_m(x,y)$

and

(1.b) $\quad |f_\delta(t,x,a) - f_\delta(t,x,b)| \leq \rho(a,b).$

By an obvious relation:

$$|\limsup_{\delta\downarrow 0} \varphi(\delta) - \limsup_{\delta\downarrow 0} \psi(\delta)| \leq \limsup_{\delta\downarrow 0}|\varphi(\delta) - \psi(\delta)|,$$

we have

(1.c) $\quad |f(f,x,a) - f(t,x,b)| \leq \rho(a,b)$

by (1.b). Using

$$\rho(x(s),a) + \rho(x(s),b) \geq \rho(a,b),$$

we have $\rho(a,b) \leq f_\delta(t,x,a) + f_\delta(t,x,b)$ and so

(2) $\quad \rho(a,b) \leq f(t,x,a) + f(t,x,b).$

$f_\delta(t,x,a)$ is continuous and so measurable $\mathcal{B}(I \times M)$ in $(t,x) \in I \times M$ for every $\delta > 0$ and every $a \in R^1$ by virtue of (1.a). Since $f_\delta(t,x,a)$ is obviously continuous in $\epsilon > 0$ for (t,x,a) fixed, we have

$$f(t,x,a) = \lim_{n\to\infty} \sup_{\substack{\propto\,\propto\frac{1}{n} \\ \delta \text{ rational}}} f_\delta(t,x,a).$$

This implies that $f(t,x,a)$ is also measurable $\mathcal{B}(I \times M)$ in $(t,x) \in I \times M$ for a fixed.

By (1.c) $f(t,x,a)$ is continuous in. $a \in R$ for (f,x) fixed. Therefore we have

$$f(t,x) = \inf_k f(t,x,a_k),$$

where $\{a_k\}$ is a countable dense subset in R^1. This implies that $f(t,x)$ is measurable $\mathcal{B}(I \times M)$.

Now we will show

(3) $$(t,x) \in \mathcal{D}(e) \Longleftrightarrow t \in \mathcal{D}(x) \Longleftrightarrow f(t,x) = 0.$$

The first equivalence is obvious by the definition. Suppose $t \in \mathcal{D}(x)$. Then $f(t,x,x(t)) = 0$ by $x \in M$ and so we have $f(t,x) = 0$. Suppose conversely that $f(t,x) = 0$. Then we have a sequence $\{b_k\}$ such that $f(t,x,b_k) \to 0$ as $k \to \infty$. By (2) we have

$$\lim_{k,h\to\infty} \rho(b_k,b_h) \leq \lim_k f(t,x,b_k) + \lim_h f(t(t,x,b_h) = 0.$$

Therefore we have $b \in R^1$ such that $\rho(b_k,b) \to 0$ as $k \to \infty$. Using (1.c) we obtain

$$|f(t,x,b_k) - f(t,x,b)| \leq \rho(b_k,b) \to 0 \quad (k \to \infty).$$

Since $f(t,x,b_k) \to 0$ $(k \to \infty)$, we have $f(t,x,b) = 0$. This implies $t \in \mathcal{D}(x)$ and $b = x(t)$ by the second part of Theorem 2 and the definition of M. This completes the proof of (3).

It remains only to prove the measurability of $\epsilon:\mathcal{D}(\epsilon) \to R^1$. Let $\{a_k\}$ be a dense sequence in R^1. Define $\tilde{\epsilon}_n(t,x)$ to be the first a_k for which

$$f(t,x,a_k) < f(t,x) + 1/n.$$

Since we have above proved $f(t,x) = \inf_k f(t,x,a_k)$, $\tilde{\epsilon}_n(t,x)$ is well defined for every (t,x). By the measurability of $f(t,x,a)$ in (t,x), it is easy to see that $\tilde{\epsilon}_n(t,x)$ is also measurable $\mathcal{B}(I \times M)$ in (t,x). Let ϵ_n be the restriction of $\tilde{\epsilon}_n$ to $\mathcal{D}(\epsilon)$. Then $\epsilon_n : I \times M \to R^1$ is measurable $\mathcal{D}(\epsilon) \cap \mathcal{B}(I \times M)/\mathcal{B}(R^1)$. If $(t,x) \in \mathcal{D}(\epsilon)$, then $t \in \mathcal{D}(x)$ and $f(t,x) = 0$, so that

$$f(t,x,\epsilon_n(t,x)) < 1/n \quad \text{and} \quad f(t,x,x(t)) = 0 .$$

Using (2) we have

$$\rho(\epsilon_n(t,x),x(t)) < f(t,x,\epsilon_n(t,x)) + f(t,x,x(t))$$

$$\to 0 \qquad (n \to \infty),$$

i.e.

$$\lim_n \epsilon_n(t,x) = x(t) = \epsilon(t,x).$$

This proves that $\epsilon : I \times M \to R^1$ is also measurable $\mathcal{D}(\epsilon) \cap \mathcal{B}(I \times M)/\mathcal{B}(R^1)$.

Generalization. As in the case of C and D, the notion of M is extended to the space $M_E(I)$ of canonical measurable functions with values in a Polish space E. No change is necessary for such an extension. The only one point to be mentioned is as follows.

Let ρ be a metric in E bounded by 1 which determines the topology in E. Then the metric ρ_m in $M_E(I)$ is defined by

$$\rho_m(x,y) = \int_I \rho(x(t),y(t))dt.$$

It should be noted that the ρ_m-topology is independent of the choice of ρ. Suppose that we have two metrics ρ' and ρ''. Then $\rho'_m(x_n,x) \to 0$ implies $\rho''_m(x_n,x) \to 0$. Suppose that it is not the case. Then we have a subsequence $\{y_n\}$ of $\{x_n\}$ such that $\rho'_m(y_n,x) \to 0$ and $\rho''_m(y_n,x) > c > 0$. By taking a subsequence again, we can assume that $\rho'_m(y_n,x) < 2^{-n}$, $n = 1,2,\dots$. Then

$$\int_I \sum_n \rho'(y_n(t),x(t))dt \leq \sum_n \rho'_m(y,x) \leq 1$$

and so $\rho'(y_n(t),x(t)) \to 0$ a.e. on I.
This implies that $\rho''(y_n(t),x(t)) \to 0$ a.e. on I. By the bounded convergence theorem we have

$$\rho''_m(y_n,x) = \int_I \rho''(y_n(t),x(t))dt \to 0,$$

in contradiction with $\rho''_m(y_n,x) > c$.

If we take a complete metric ρ in E bounded by 1, then ρ_m is a complete metric in $M_E(I)$.

In case $I = [0,\infty)$ or $(-\infty,\infty)$, we need not make special consideration about the right end point as we did for $I = [0,1]$. The ρ_m-metric is to be defined by

$$\rho_m(x,y) = \int_I \rho(x(t),y(t)) \frac{dt}{1+t^2} .$$

1.8 The space $\mathcal{D}'$ of distributions

Let $C^\infty = C^\infty(R^1)$ be the space of all functions: $R^1 \to R^1$ that are continuously differentiable infinitely many times. The support $S(\varphi)$ of $\varphi \in C^\infty$ is defined to be the closure of the set $\{t : \varphi(t) \neq 0\}$. The space of all functions $\in C^\infty(R^1)$ with compact support is denoted by $\mathcal{D} = \mathcal{D}(R^1)$. $\mathcal{D}$ is obviously a linear space with the usual scalar multiplication and the usual addition. We will write $\varphi_n \twoheadrightarrow \varphi$ $(\varphi_1, \varphi_2, \ldots, \varphi \in \mathcal{D})$ if

(i) $\varphi_n^{(k)}(t) \to \varphi^{(k)}(t)$ uniformly in t for every $k = 0, 1, 2, \ldots$

and

(ii) the closure of $\bigcup_n S(\varphi_n)$ is compact.

A real linear functional x on $\mathcal{D}$ is called a distribution in R^1 if $\varphi_n \twoheadrightarrow 0$ implies $x(\varphi_n) \to 0$. The space of all distributions on R^1 is denoted by $\mathcal{D}' = \mathcal{D}(R^1)$. $\mathcal{D}'$ is clearly regarded as a linear space in an obvious way.

Let $\mathcal{D}(a)$ be the space of all $\varphi \in \mathcal{D}$ with support $\subset [-a, a]$. $\mathcal{D}(a)$ is clearly a linear subspace on $\mathcal{D}$. For $\varphi_1, \varphi_2, \ldots, \varphi \in \mathcal{D}(a)$, $\varphi_n \twoheadrightarrow \varphi$ if and only if the condition (i) is satisfied, because (ii) is automatic in this case. A real linear functional x on $\mathcal{D}(a)$ is called a distribution on $[-a, a]$ if $\varphi_n \twoheadrightarrow 0$ implies $x(\varphi_n) \to 0$. The space of all distributions on $[-a, a]$ is denoted by $\mathcal{D}(a)'$. $\mathcal{D}(a)'$ is also regarded as a linear subspace in an obvious way. If $x \in \mathcal{D}'$, then the restriction $x|\mathcal{D}(a)$ of x to $\mathcal{D}(a)$ belongs to $\mathcal{D}(a)'$, but not every distribution on $[-a, a]$ can be extended to a distribution in $(-\infty, \infty)$. $\mathcal{D}[0, \infty)$ and $\mathcal{D}[0, \infty)'$ can be defined similarly.

The Kolmogorov σ-algebra $\mathcal{B}_K(\mathcal{D}')$ is defined to be the σ-algebra on $\mathcal{D}'$ generated by the sets

$$\{x \in \mathcal{D}' : x(\varphi) < c\}, \ \varphi \in \mathcal{D}, \ c \in R^1.$$

$\mathcal{B}_K(\mathcal{D}(a)')$ and $\mathcal{B}_K(\mathcal{D}[0,\infty)')$ are defined similarly. The purpose of this section is to prove that $\mathcal{D}'$, $\mathcal{D}(a)'$ and $\mathcal{D}[0,\infty)'$ with the Kolmogorov σ-algebras are standard Borel spaces.

The <u>k-th inner product</u> $(\ , \ )_k$ in $\mathcal{D}(a)$ is defined by

$$(\varphi,\psi)_k = \int_{-\infty}^{\infty} \varphi^{(k)}(t)\psi^{(k)}(t)dt.$$

It is easy to check that (i) $(\varphi,\psi)_k$ is bilinear in (φ,ψ) and (ii) $(\varphi,\varphi)_k > 0$ $(\varphi \neq 0)$ and $(0,0)_k = 0$. $\mathcal{D}(a)$ is obviously a pre-Hilbert space with the k-th inner product for each k. The norm $\|\varphi\|_k = \sqrt{(\varphi,\varphi)_k}$ is called the <u>k-th norm</u>. The space $\mathcal{D}(a)$ is separable, because $\mathcal{D}(a)$ is isomorphic with a subspace of $L^2(R^1)$ by the map $\varphi \to \varphi^{(k)}$.

The space of all real linear functional on $\mathcal{D}(a)$ continuous with respect to the k-th norm topology is denoted by $\mathcal{D}(a)'_k$. Since $\varphi_n \twoheadrightarrow 0$ implies $\|\varphi_n\|_k \to 0$, $\mathcal{D}(a)'_k$ is clearly a linear subspace of $\mathcal{D}(a)'$. Using the Schwartz inequality, we can prove that

(1) $$\|\varphi\|_k \leq 2a\|\varphi\|_{k+1} \quad \text{for} \quad \varphi \in \mathcal{D}(a).$$

This implies that

(2) $$\mathcal{D}(a)'_1 \subset \mathcal{D}(a)'_2 \subset \dots .$$

Define the <u>(-k)-th norm</u> $\| x \|_{-k}$ of $x \in \mathcal{D}(a)'$ by

(3) $$\| x \|_{-k} = \sup\{ |x(\varphi)| : \| \varphi \|_k \leq 1 \}.$$

$\| x \|_{-k}$ may be ∞, but $\| x+y \|_{-k} \leq \| x \|_{-k} + \| y \|_{-k}$ and $\| cx \|_{-k} = |c| \ \| x \|_{-k}$. It is also true that $|x(\varphi)| \leq \| x \|_{-k} \| \varphi \|_k$. Since $\mathcal{D}(a)$ with the $\| \ \|_k$-topology is separable, we have a countable dense set $\{\varphi_n\}$ in $\mathcal{D}(a)$ and $\| x \|_{-k}$ is expressed as

(3') $$\| x \|_{-k} = \sup_{n : \| \varphi_n \|_k \leq 1} |x(\varphi_n)|$$

<u>Theorem 1</u>. $\mathcal{D}(a)'_k = \{ x \in \mathcal{D}(a)' : \| x \|_{-k} < \infty \}$.

<u>Proof</u>. Suppose that $x \in \mathcal{D}(a)'$. Then x is a linear functional in $\mathcal{D}(a)$. Therefore x is continuous with respect to the $\| \ \|_k$-topology if and only if $\| x \|_{-k} < \infty$.

<u>Theorem 2</u>. $\mathcal{D}(a)' = \bigcup_k \mathcal{D}(a)'_k$.

<u>Proof</u>. It is enough to prove that every $x \in \mathcal{D}(a)'$ belongs to some $\mathcal{D}(a)'_k$. Suppose that $x \notin \mathcal{D}(a)'_k$ for every k. Then we can find $\{\varphi_k\} \subset \mathcal{D}(a)$ such that

(4) $$|x(\varphi_k)| > (2a)^{2k} \| \varphi_k \|_k, \quad k = 0, 1, 2, \ldots .$$

It is obvious that $\| \varphi_k \| \neq 0$. Set

$$\psi_k = \frac{1}{(2a)^{2k} \| \varphi_k \|_k} \cdot \varphi_k.$$

Then for $k \geq n$ we have

$$\|\psi_k\|_n = \frac{\|\varphi_k\|_n}{(2a)^{2k}\|\varphi_k\|_k} \leq \frac{(2a)^{k-n}}{(2a)^{2k}} = (2a)^{-k-n} \quad \text{(by (1))}$$

$$\to 0 \quad (k \to \infty), \; n = 1,2,\ldots$$

and so

$$|\psi_k^{(n)}(t)| \leq \int_{-a}^{t} |\psi_k^{(n+1)}(s)|ds \leq \left(\int_{-a}^{t} \psi_k^{(n+1)}(s)^2 ds\right)^{\frac{1}{2}}(t+a)^{\frac{1}{2}}$$

$$\leq \|\psi_k\|_{n+1}(2a)^{\frac{1}{2}} \to 0 \quad (k \to \infty) \quad \text{for } t \in [-a,a]$$

$$\psi_k^{(n)}(t) = 0 \qquad\qquad \text{for } t \notin [-a,a].$$

Therefore $\psi_k^{(n)}(t) \to 0$ uniformly in t for every n. Since $x \in \mathcal{D}(a)'$, $x(\psi_k) \to 0$. But

$$|x(\psi_k)| = \frac{1}{(2a)^{2k}\|\varphi_k\|_k} \; |x(\varphi_k)|$$

$$> 1 \qquad \text{(by (4))}.$$

This is a contradiction.

<u>Theorem 3</u>. There exists an inner product $(\, , \,)_{-k}$ in $\mathcal{D}(a)'_k$ such that $(x,x)_{-k} = \|x\|^2_{-k}$. The space $\mathcal{D}(a)'_k$ with this inner product is a separable Hilbert space.

<u>Proof</u>. Since $\mathcal{D}(a)$ is a separable pre-Hilbert space with $(\ ,\ )_k$, we can find a complete orthonormal system $\{\varphi_n\}$ by the Schmidt orthogonalization method. Every element $\varphi \in \mathcal{D}(a)$ can be expressed as

$$\varphi = \sum_{n=1}^{\infty} a_n\varphi_n \qquad (\text{convergence in } \|\ \|_k)$$

where $\sum_n a_n^2 = \|\varphi\|_k^2 < \infty$. Observing

$$|x(\varphi)|^2 = (\sum_n a_n x(\varphi_n))^2 \leq \sum_n a_n^2 \sum_n x(\varphi_n)^2 = \|\varphi\|_k^2 \sum_n x(\varphi_n)^2$$

for $x \in \mathcal{D}(a)'_k$, we have

$$\|x\|_{-k}^2 \leq \sum_n x(\varphi_n)^2.$$

On the other hand, we have

$$|x(\sum_1^N a_n\varphi_n)|^2 \leq \|x\|_{-k}^2 \; \|\sum_1^N a_n\varphi_n\|^2 = \|x\|_{-k}^2 \sum_1^N a_n^2 .$$

Setting $a_n = x(\varphi_n)$ we have

$$(\sum_1^N x(\varphi_n)^2)^2 \leq \|x\|_{-k}^2 \sum_1^N x(\varphi_n)^2$$

and so

$$\sum_1^N x(\varphi_n)^2 \leq \|x\|_{-k}^2.$$

Letting $N \uparrow \infty$ we have

$$\sum_1^{\infty} x(\varphi_n)^2 \leq \|x\|_{-k}^2.$$

Therefore we have

(5) $$\sum_1^\infty x(\varphi_n)^2 = \|x\|_{-k}^2.$$

Define $(x,y)_{-k}$ by

(6) $$(x,y)_{-k} = \sum_1^\infty x(\varphi_n)y(\varphi_n).$$

It is easy to see that this satisfies all conditions of an inner product and that $(x,x)_{-k} = \|x\|_{-k}^2$.

Since the space $\mathcal{D}(a)_k'$ with the inner product $(\ ,\ )_{-k}$ is isomorphic with ℓ^2 by the map

$$f: x \to (x(\varphi_1), x(\varphi_2), \dots),$$

it is a separable Hilbert space.

<u>Theorem 4</u>. The topological σ-algebra $\mathcal{B}(\mathcal{D}(a)_k')$ is identical with the trace of $\mathcal{B}_K(\mathcal{D}(a)')$ to $\mathcal{D}(a)_k'$.

<u>Proof</u>. Since for each $\varphi \in \mathcal{D}(a)$ the map $x \to x(\varphi)$ is continuous in $x \in \mathcal{D}(a)_k'$ by

$$|x(\varphi) - y(\varphi)| \leq \|x-y\|_{-k}\|\varphi\|_k,$$

the set $\{x \in \mathcal{D}(a)' : x(\varphi) < c\}$ is open in $\mathcal{D}(a)_k'$. This implies that

$$\mathcal{D}(a)_k' \cap \mathcal{B}_k(\mathcal{D}(a)') \subset \mathcal{B}(\mathcal{D}(a)_k').$$

By (3') we have

$$\| x-x_0 \|_{-k} = \sup_n |x(\varphi_n) - x_0(\varphi_n)|$$

and so the ball $\{x \in \mathcal{D}(a)'_k : \| x-x_0 \|_{-k} \leq r\}$ belongs to $\mathcal{D}(a)'_k \cap \mathcal{B}_k(\mathcal{D}(a)')$. Since $\mathcal{B}(\mathcal{D}(a)'_k)$ is generated by such balls, we have

$$\mathcal{B}(\mathcal{D}(a)'_k) \subset \mathcal{D}(a)'_k \cap \mathcal{B}_k(\mathcal{D}(a)').$$

Theorem 5. The Borel space $(\mathcal{D}(a)', \mathcal{B}_k(\mathcal{D}(a)'))$ is standard.

Proof. $\mathcal{D}(a)'$ is the union of $\mathcal{D}(a)'_k$, $k = 1,2,\ldots$. Since $\| x \|_{-k}$ is measurable $\mathcal{B}_K(\mathcal{D}(a)')$ by (3'), $\mathcal{D}(a)'_k \in \mathcal{B}_K(\mathcal{D}(a)')$ by Theorem 1. Since $\mathcal{D}(a)'_k$ is a separable Hilbert space, it is Polish and so $(\mathcal{D}(a)'_k, \mathcal{B}(\mathcal{D}(a)'_k))$ is standard. But $\mathcal{B}(\mathcal{D}(a)'_k)$ is identical with the trace σ-algebra of $\mathcal{B}_K(\mathcal{D}(a)')$ on $\mathcal{D}(a)'_k$. Now we can use 1.3 Theorem 5 to complete the proof.

Theorem 6. The Borel space $(\mathcal{D}', \mathcal{B}_K(\mathcal{D}'))$ is standard.

Proof. Consider the product space $P = \prod_{n=1}^{\infty} \mathcal{D}(n)'$ and denote the product σ-algebra $\otimes_n \mathcal{B}_K(\mathcal{D}(n)')$ by $\mathcal{P}$. For every $x \in \mathcal{D}'$ the restriction $x_n = x/\mathcal{D}(n)$ belongs to $\mathcal{D}(n)'$ and so

$$f(x) \equiv (x_1, x_2, \ldots) \in P.$$

The map f is obviously 1-1 from $\mathcal{D}'$ into P because every $\varphi \in \mathcal{D}$ belongs to some $\mathcal{D}(n)$. Let $\xi = (\xi_1, \xi_2, \ldots)$ be an arbitrary element in P. Then it is easy to see that $\xi \in f(\mathcal{D}')$ if and only if

$$\xi_n = \xi_m/\mathcal{D}(n) \quad \text{for} \quad n < m.$$

Let π_n denote the projection from P onto $\mathcal{D}(n)'$ and α_{nm} the restriction map from $\mathcal{D}(m)'$ into $\mathcal{D}(n)'$, i.e.

$$\alpha_{nm}(\xi_m) = \xi_m/\mathcal{D}(n), \ \xi_m \in \mathcal{D}(m), \ n < m.$$

Then

$$f(\mathcal{D}') = \{\xi \in P: (\alpha_{nm} \circ \pi_m)(\xi) = \pi_n(\xi) \text{ for } n < m\}.$$

$\pi_m: P \to \mathcal{D}(n)'$ is clearly measurable $\mathcal{P}/\mathcal{B}_K(\mathcal{D}(n)')$ and $\alpha_{nm}: \mathcal{D}(m)' \to \mathcal{D}(n)'$ is also measurable $\mathcal{B}_K(\mathcal{D}(m)')/\mathcal{B}_K(\mathcal{D}(n)')$ by the definition. Therefore $\alpha_{nm} \cdot \pi_m$ and π_n are both measurable $\mathcal{P}/\mathcal{B}_K(\mathcal{D}(n)')$. Since $(\mathcal{D}(n)', \mathcal{B}_K(\mathcal{D}(n)'))$ is standard, the set

$$\{\xi \in P: (\alpha_{nm} \circ \pi_m)(\xi) = \pi_n(\xi)\}$$

belongs to $\mathcal{P}$ by <u>1.3</u> Theorem 6. Therefore $f(\mathcal{D}') \in \mathcal{P}$.

As $(\mathcal{D}(n)', \mathcal{B}_K(\mathcal{D}(n)'))$ is standard by 1.3 Theorem 6 so is $(P, \mathcal{P})$ by <u>1.3</u> Theorem 4. Therefore the space $f(\mathcal{D}')$ with the trace σ-algebra $f(\mathcal{D}') \cap \mathcal{P}$ is also standard by $f(\mathcal{D}') \in \mathcal{P}$. It is easy to see that $(f(\mathcal{D}'), f(\mathcal{D}') \cap \mathcal{P})$ is Borel isomorphic with $(\mathcal{D}', \mathcal{B}_K(\mathcal{D}'))$. This completes the proof.

In the same way as above we can prove the following.

<u>Theorem 7</u>. The Borel space $(\mathcal{D}'[0,\infty), \mathcal{B}_K(\mathcal{D}'[0,\infty)))$ is standard.

Chapter 2. Basic Concepts in Probability Theory

2.1 Probability measures.

Let S be an arbitrary space. A map μ from a σ-algebra $\mathfrak{M} = \mathfrak{M}(\mu)$ on S into $[0,\infty]$ is called a measure on S or on $\mathfrak{M}$ if it satisfies the following conditions:

(μ.1) $\mu(\emptyset) = 0$,

(μ.2) (σ-additive) $\mu(\cup_n A_n) = \Sigma_n \mu(A_n)$ for disjoint $A_1, A_2, \ldots, \in \mathfrak{M}$.

μ is called a probability measure if $\mu(S) = 1$. A space S endowed with a measure μ on S is called a measure space (S,μ). If μ is a probability measure, (S,μ) is called a probability space.

We assume the reader to be familiar with the general theory of measures and integrals. In this section, we will mention some fundamental facts on probability measures which may not be found in a standard text book on measure theory.

Suppose that S is a Borel space endowed with a σ-algebra $\mathcal{S}$ on S. A probability measure μ on S is called $\mathcal{S}$-regular or regular on $S = (S,\mathcal{S})$ if it satisfies the following conditions:

(R.1) μ is complete,

(R.2) $\mathfrak{M}(\mu) \supset \mathcal{S}$,

(R.3) for every $A \in \mathfrak{M}(\mu)$ we have $B \in \mathcal{S}$ such that $B \subset A$ and $\mu(A-B) = 0$.

In other words a probability measure on S is called regular on $(S, \mathcal{S})$ if it is the Lebesgue extension of a probability measure on $\mathscr{S}$. Two (complete) regular probability measures on $(S,\mathscr{S})$ are identical if they coincide on $\mathscr{S}$.

Let S be a Hausdorff topological space. S is regarded as a Borel space with the topological σ-algebra $\mathscr{B}(S)$. Therefore we can define regular measures on S. We will define two stronger notions of regularity, F-regularity and K-regularity. A probability measure μ on S is called <u>F-regular</u> (or <u>K-regular</u>) if it satisfies (R.1), (R.2) and (R.3): "for every $A \in \mathfrak{M}(\mu)$ and every $\epsilon > 0$ we can find a closed (or compact) $C = C(A,\epsilon) \subset A$ such that $\mu(A - C) < \epsilon$". Every K-regular probability measure is F-regular and every F-regular probability measure is regular.

<u>Theorem 1</u>. Suppose that every open subset of S is an F_σ-set. (For example, every metrizable space S has this property.) Then every regular probability measure on S is F-regular.

<u>Proof</u>. Let $\mathfrak{A}$ be the class of all subsets $A \in \mathscr{B}(S)$ such that for every $\epsilon > 0$ we can find an open set $G \supset A$ and a closed set $F \subset A$ with $\mu(G-F) < \epsilon$. It is easy to check that $\mathfrak{A}$ is a σ-algebra. Since every open set belongs to $\mathfrak{A}$ by the assumption, $\mathfrak{A}$ includes $\mathscr{B}(S)$. This completes the proof.

<u>Theorem 2</u>. (<u>Prohorov's theorem</u>). Suppose that S is a Polish space. Then every regular probability measure μ on S is K-regular.

<u>Proof</u>. First we will construct a compact set $K = K(\epsilon)$ for $\epsilon > 0$ such that $\mu(S-K) < \epsilon/2$. Let ρ be a complete metric

in S determining the topology in S. Let $\{a_1, a_2, \ldots\}$ be a countable dense set and denote by B_{mn} the closed ball with center a_m and radius $1/n$. Then we have

$$S = \cup_m B_{mn}, \quad n = 1,2, \ldots .$$

Therefore we can find $M = M(n)$ for every $n = 1,2, \ldots$ such that

$$\mu(S - \bigcup_{m=1}^{M(n)} B_{mn}) < 2^{-n-1}\epsilon.$$

Set

$$F_n = \bigcup_{m=1}^{M(n)} B_{mn} \quad \text{and} \quad K = \bigcap_n F_n.$$

Then

$$\mu(S-K) = \mu(\bigcup_n (S - F_n)) \leq \sum_n \mu(S - F_n) < \epsilon/2.$$

Since every B_{mn} is closed, every F_n is closed and so is K. Since K is covered by B_{mn}, $m = 1,2,\ldots, M(n)$ for every n, K is totally bounded. Therefore K is compact as a totally bounded closed subset of a complete metric space (S,ρ).

Let A be an arbitrary Borel set in the class $\mathfrak{M}(\mu)$. Since μ is F-regular by Theorem 1, we can find a closed set $F \subset A$ such that $\mu(A - F) < \epsilon/2$. Let K' denote the intersection of F with the compact set K constructed above. Then K' is a compact subset of A and we have

$$\mu(A-K') \leq \mu((A-F) \cup (A-K))$$
$$\leq \mu(A-F) + \mu(A-K)$$
$$\leq \mu(A-F) + \mu(S-K) < \epsilon.$$

This completes the proof.

As an example of this theorem we have that every regular probability measure on $R^n (n = 1,2,\ldots,\infty)$ is K-regular.

Let f be a map from S into T and μ a probability measure on S. Then the class $\mathcal{N}$ of all subsets B of T such that $f^{-1}(B) \in \mathfrak{M}(\mu)$ is a σ-algebra on T and

$$\nu(B) = \mu(f^{-1}(B)), \quad B \in \mathcal{N},$$

defines a probability measure on T with $\mathfrak{M}(\nu) = \mathcal{N}$. The probability measure ν is called the <u>image measure</u> of μ by the map f and denoted by $f \cdot \mu$ or $\mu \cdot f^{-1}$. If g is a map from T into U, then we have

$$(g \circ f) \cdot \mu = g \cdot (f \cdot \mu).$$

Let ν be the image measure of a measure μ on S by a map $f: S \to T$, g a ν-measurable real or complex function on T and B a ν-measurable subset of T. Then $g \circ f$ and $f^{-1}(B)$ are μ-measurable and we have the following <u>transformation formula</u>:

$$\int_{f^{-1}(B)} (g \circ f)(x) \mu(dx) = \int_B g(y) \nu(dy).$$

The equality means that if one of the two integrals exists, then the other exists and they are equal.

Let (S, μ) be a probability space and $(T, \mathcal{T})$ a Borel space. A map $f: S \to T$ is called <u>μ-measurable</u> if it is measurable $\mathfrak{M}(\mu) | \mathcal{T}$. The image measure $f \cdot \mu$ is not regular on $(T, \mathcal{T})$ in general.

Suppose that $(S, \mathcal{S})$ and $(T, \mathcal{T})$ are Borel spaces and that μ is a regular probability measure on $(S, \mathcal{S})$. If $f: S \to T$ is μ-measurable i.e. measurable $\mathfrak{M}(\mu) | \mathcal{T}$, then the image measure $f \cdot \mu$ is a complete probability measure on T with $\mathfrak{M}(f \cdot \mu) \supset \mathcal{T}$.

Even in this case $f \cdot \mu$ is not always regular; see the example at the end of this section. The following theorem plays an important role in this connection.

Theorem 3. Suppose that $(S, \mathcal{S})$ and $(T, \mathcal{T})$ are standard Borel spaces and that μ is a regular probability measure on $(S, \mathcal{S})$. If $f: S \to T$ is μ-measurable, then the image measure $f \cdot \mu$ is regular.

Proof. Since the regularity of measures is invariant under Borel isomorphism, we can assume that S and T are Borel subsets of R^1 endowed with the trace σ-algebras $S \cap \mathcal{B}^1$ and $T \cap \mathcal{B}^1$ respectively. Define a probability measure ν on R^1 by

$$\nu(E) = \mu(E \cap S)$$

for all $E \subset R^1$ such that $E \cap S \in \mathcal{M}(\mu)$. Since S is a Borel subset of R^1, every set $B \in S \cap \mathcal{B}^1$ is also a Borel subset of R^1. Noticing this fact we can easily check that ν is a regular probability measure on R^1 with $\mathcal{M}(\nu) \supset \mathcal{M}(\mu)$. Extend f to a map $g: R^1 \to R^1$ by

$$g(x) = \begin{cases} f(x), & x \in S \\ 0, & x \in R^1 - S. \end{cases}$$

Noticing that $\nu(R^1 - S) = 0$, we can easily see that g is a ν-measurable map from R^1 into itself. Therefore for every ν-measurable set $E \subset R^1$ and every $\epsilon > 0$ we can find a compact subset K of E such that

(i) $\nu(E-K) < \epsilon$ and

(ii) the restriction $g|K$ is continuous (Lusin's Theorem).

Keeping this observation in mind we will prove that the image measure $\theta = f\cdot\mu$ is a regular probability measure on $(T, \mathcal{T})$.

It is obvious that θ is a complete probability measure on T with $\mathcal{M}(\theta) \supset \mathcal{T} = T \cap \mathcal{B}^1$. For completion of the proof it is enough to show that for every $A \in \mathcal{M}(\theta)$ and every $\epsilon > 0$ we can construct a compact subset H of A with $\theta(A-H) < \epsilon$. Since $A \in \mathcal{M}(\theta)$, we have $f^{-1}(A) \in \mathcal{M}(\mu) \subset \mathcal{M}(\nu)$. By Lusin's Theorem we have a compact set $K \subset f^{-1}(A)$ such that

(i) $\mu(f^{-1}(A)-K) < \epsilon$ and

(ii) the restriction $g_K = g|K$ is continuous. Set

$$H = f(K).$$

Then it is obvious that $H \subset A$ by $K \subset f^{-1}(A)$. Since $K \subset f^{-1}(A) \subset S$, we have

$$H = g(K) = g_K(K).$$

H is compact as a continuous image of a compact set K. Since $H = f(K)$, we have $f^{-1}(H) \supset K$. Therefore

$$\theta(A-H) = \mu(f^{-1}(A)-f^{-1}(H)) \leq \mu(f^{-1}(A)-K) < \epsilon.$$

This completes the proof of regularity of θ.

A standard Borel space $(S, \mathcal{S})$ endowed with a regular probability measure μ on $(S, \mathcal{S})$ is called a <u>standard probability space</u> $(S, \mathcal{S}, \mu)$. This definition is slightly more restricted than that of G. Mackey which reads as follows:

A Borel space $(S, \mathcal{S})$ endowed with a regular probability measure is called a standard probability space if we have a set $S_1 \in \mathcal{S}$ with $\mu(S_1) = 1$ such that the Borel space $(S_1, S_1 \cap \mathcal{S})$ is standard. In this book we will not take this definition.

<u>Example</u>. Let λ be the Lebesgue measure on $I = [0,1]$. Take a subset S of I such that (i) $S \notin \mathcal{M}(\lambda)$ and (ii) $\underline{\lambda}(I-S) = 0$, where $\underline{\lambda}$ denotes the inner Lebesgue measure. The famous example of non-measurable sets satisfies these conditions. S is a Borel space with the trace σ-algebra $\mathcal{S} = S \cap \mathcal{B}^1$. Define μ by

$$\mu(A) = \lambda(M)$$

for every set $A \subset S$ of the form $A = M \cap S$, $M \in \mathcal{M}(\lambda)$. If $M_1 \cap S = M_2 \cap S$, then $(M_1 \ominus M_2) \cap S = \emptyset$. Therefore

$$\lambda(M_1 \ominus M_2) \leq \underline{\lambda}(I-S) = 0.$$

This shows that μ is well-defined. It is easy to check that μ is a regular probability measure on $(S, \mathcal{S})$. Let f be the identity map from S into I. If $B \in \mathcal{B}(I)$, then $B \in \mathcal{M}(\lambda)$ and so

$$f^{-1}(B) = S \cap B \in \mathcal{M}(\mu).$$

This implies that f is μ-measurable i.e., measurable $\mathcal{M}(\mu)|\mathcal{B}(I)$. Nevertheless the image measure $\nu = f\cdot\mu$ is not regular. If $B \in \mathcal{B}(I)$, then $f^{-1}(B) = S \cap B \in \mathcal{M}(\mu)$. Therefore

$$B \in \mathcal{M}(\nu) \quad \text{and} \quad \nu(B) = \mu(f^{-1}(B)) = \mu(S \cap B) = \lambda(B).$$

This shows that $\mathcal{B}(I) \subset \mathcal{M}(\nu)$ and $\nu = \lambda$ on $\mathcal{B}(I)$. Suppose that ν is regular. Then $\nu = \lambda$ and so $\mathcal{M}(\nu) = \mathcal{M}(\lambda)$. Since $f^{-1}(S) = S = S \cap I \subset \mathcal{M}(\mu)$, we have

$$S \in \mathcal{M}(\nu) = \mathcal{M}(\lambda),$$

which contradicts $S \notin \mathcal{M}(\lambda)$. Therefore ν is not regular.

 (~~add this between p.9 and p.10 in Chapter 2~~)

Let μ be a probability measure and A_λ, $\lambda \in \Lambda$ be a system of μ-measurable sets such that $\cup_{\lambda \in \Lambda} A_\lambda$ is also μ-measurable. Then it is obvious that

$$\mu(\bigcup_{\lambda \in \Lambda} A_\lambda) \geq \sup_M \mu(\bigcup_{\lambda \in M} A_\lambda)$$

where M moves over all finite subsets of Λ. If Λ is countable, we have only a countable number of possible choices of M and so the equality holds. If Λ is not countable, the left-hand side may be larger than the right one. But we have the following theorem.

<u>Theorem 4</u>. ~~If~~ Let μ ~~is~~ be a K-regular probability measure on a ~~compact~~ Hausdorff topological space S.

(a) If G_λ is open for every $\lambda \in \Lambda$, then

$$\mu(\bigcup_{\lambda \in \Lambda} G_\lambda) = \sup_M \mu(\bigcup_{\lambda \in M} G_\lambda)$$

(b) If F_λ is closed for every $\lambda \in \Lambda$, then

$$\mu(\bigcap_{\lambda \in \Lambda} F_\lambda) = \inf_M \mu(\bigcap_{\lambda \in M} F_\lambda).$$

In both cases M moves over all finite subsets of Λ.

<u>Proof</u>. Set $G = \cup_{\lambda \in \Lambda} G_\lambda$. This is obviously open and so μ-measurable. Since μ is K-regular, we have a compact set $K \subset G$ for every $\epsilon > 0$ such that $\mu(G) < \mu(K) + \epsilon$. By the covering theorem we have a finite subset M of Λ such that

$$\bigcup_{\lambda \in M} G_\lambda \supset K.$$

Therefore

$$\mu(\bigcup_{\lambda \in M} G_\lambda) \geq \mu(K) > \mu(G) - \epsilon.$$

This implies that the right hand side of the equation in (a) is no less than the left one. The opposite inequality is obvious. This completes the proof of (a). To prove (b), apply (a) to the open sets F_λ^c, $\lambda \in \Lambda$.

As we explained before, regularity is not inherited by image measures in general but we have the following.

Theorem 5. Let μ be a K-regular probability measure on a Hausdorff topological space S and f a continuous map from S into another Hausdorff space T. Then the image measure $\nu = f\cdot\mu$ is a K-regular probability measure on T.

Proof. It is obvious that ν is a complete probability measure on T such that $\mathfrak{M}(\nu) \supset \mathcal{B}(T)$. For completion of the proof it is enough to show that for every $B \in \mathfrak{M}(\nu)$ and every $\epsilon > 0$ we can find a compact subset K of B such that

$$\nu(B - K) < \epsilon.$$

By the definition of the image measure we have

$$A = f^{-1}(B) \in \mathfrak{M}(\mu) \quad \text{and} \quad \mu(A) = \nu(B).$$

Since μ is K-regular, we have a compact subset H of A such that $\mu(A - H) < \epsilon$. Since f is continuous, the image $K = f(H)$ is compact and

$$K \subseteq f(A) \subseteq B.$$

$f^{-1}(K)$ is a closed subset of A including H. Therefore

$$\nu(B - K) = \mu(f^{-1}(B) - f^{-1}(K)) \leq \mu(A - H) < \epsilon.$$

2.2 The Coincidence Theorem and the Extension Theorem

Let μ be a probability measure on a space S. To determine μ, we need not specify $\mu(A)$ for every $A \in \mathfrak{M}(\mu)$. For example, if we know the values $\mu(A)$, $\mu(B)$ and $\mu(A \cap B)$, then the value $\mu(A \cup B)$ is automatically determined by the relation:

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B).$$

It is often necessary to determine a probability measure by knowing its behavior on a certain subclass of $\mathfrak{M}(\mu)$. We have two types of theorems in this respect. The coincidence theorem gives the conditions under which two probability measures μ and ν are identical if they coincide on a subclass $\mathfrak{A}$ of $\mathfrak{M}(\mu) \cap \mathfrak{M}(\nu)$ and the extension theorem gives the conditions under which a set function defined on a class $\mathfrak{A}$ of subsets of S can be extended to a probability measure on S.

A class $\mathfrak{A}$ of subsets of S is called a multiplicative class on S if it is closed under intersection, i.e. if $A, B \in \mathfrak{A}$ implies $A \cap B \in \mathfrak{A}$. A class $\mathfrak{D}$ of subsets of S is called a Dynkin class on S if it satisfies the following conditions:

(D.1) $S \in \mathfrak{D}$,

(D.2) $\mathfrak{D}$ is closed under proper difference, i.e. $A, B \in \mathfrak{D}$ and $A \supset B$ imply $A - B \in \mathfrak{D}$,

(D.3) $\mathfrak{D}$ is closed under countable disjoint union, i.e. if $A_1, A_2, \ldots \in \mathfrak{D}$ are disjoint, then $\bigcup_n A_n \in \mathfrak{D}$.

In the same way as the σ-algebra $\sigma[\mathfrak{A}]$ generated by $\mathfrak{A}$ was defined, we can define the *Dynkin class generated by* $\mathfrak{A}$ to be the smallest Dynkin class containing $\mathfrak{A}$. This Dynkin class is denoted by $\mathfrak{D}[\mathfrak{A}]$.

Lemma 1 (Dynkin's lemma). Let $\mathfrak{A}$ be a multiplicative class. Then we have

$$\mathfrak{D}[\mathfrak{A}] = \sigma[\mathfrak{A}].$$

Proof. This lemma is similar to the monotone class theorem and is more convenient in many cases. Since a σ-algebra is also a Dynkin class, $\mathfrak{D}[\mathfrak{A}]$ is obviously included by $\sigma[\mathfrak{A}]$. We will prove the opposite inclusion. For this purpose it is enough to show that

$$(1) \qquad A, B \in \mathfrak{D}[\mathfrak{A}] \Longrightarrow A \cap B \in \mathfrak{D}[\mathfrak{A}];$$

once this is done, we can see that $\mathfrak{D}[\mathfrak{A}]$ is a σ-algebra $\supset \mathfrak{A}$ and therefore includes $\sigma[\mathfrak{A}]$. First we will prove that

$$(2) \qquad A \in \mathfrak{A},\ B \in \mathfrak{D}[\mathfrak{A}] \Longrightarrow A \cap B \in \mathfrak{D}[\mathfrak{A}].$$

Let A be an arbitrary member of $\mathfrak{A}$ and set

$$\mathfrak{D}_A = \{B \subset S : A \cap B \in \mathfrak{D}[\mathfrak{A}]\}.$$

Since $\mathfrak{A}$ is multiplicative, we have

$$B \in \mathfrak{A} \Longrightarrow A \cap B \in \mathfrak{A} \subset \mathfrak{D}[\mathfrak{A}].$$

Therefore $\mathfrak{A} \subset \mathfrak{D}_A$. As $\mathfrak{D}[\mathfrak{A}]$ is a Dynkin class, so is $\mathfrak{D}_A$. Therefore $\mathfrak{D}_A \supset \mathfrak{D}[\mathfrak{A}]$. This proves (2). Let A be an arbitrary member of $\mathfrak{D}[\mathfrak{A}]$ and define $\mathfrak{D}_A$ as above. Then $\mathfrak{D}_A$ is a Dynkin class which includes $\mathfrak{A}$ by (2). Therefore $\mathfrak{D}_A \supset \mathfrak{D}[\mathfrak{A}]$. This proves (1).

Theorem 1. (The coincidence theorem). Suppose that μ and ν are probability measures on S and that $\mathfrak{A}$ is a multiplicative class included by $\mathfrak{M}(\mu) \cap \mathfrak{M}(\nu)$. If $\mu = \nu$ on $\mathfrak{A}$, then $\mu = \nu$ on $\sigma[\mathfrak{A}]$.

Proof. Consider the class

$$\mathfrak{D} = \{A \in \mathfrak{M}(\mu) \cap \mathfrak{M}(\nu) : \mu(A) = \nu(A)\}.$$

$\mathfrak{D}$ includes $\mathfrak{A}$ by our assumption and $\mathfrak{D}$ is a Dynkin class by the properties of probability measures. Therefore $\mathfrak{D} \supset \mathfrak{D}[\mathfrak{A}]$ but $\mathfrak{D}[\mathfrak{A}] = \sigma[\mathfrak{A}]$ by Dynkin's lemma. This completes the proof.

In case S is endowed with a σ-algebra $\mathcal{S}$, we have the following as an immediate result of Theorem 1.

Theorem 2. Let μ and ν be two regular measures on $(S, \mathcal{S})$ and $\mathfrak{A}$ a multiplicative class generating $\mathcal{S}$. If $\mu = \nu$ on $\mathfrak{A}$, then $\mu = \nu$, i.e. $\mathfrak{M}(\mu) = \mathfrak{M}(\nu)$ and $\mu = \nu$ on this common domain of definition.

A class $\mathfrak{A}$ of subsets of a space S is called an algebra on S if it satisfies the following conditions:

(A.1) $S \in \mathfrak{A}$,

(A.2) $A \subset \mathfrak{A} \Longrightarrow A^c (= S - A) \in \mathfrak{A}$,

(A.3) (additive) $A, B \in \mathfrak{A} \Longrightarrow A \cup B \in \mathfrak{A}$.

A map m from an algebra $\mathfrak{A}$ on S into $[0,1]$ is called an <u>elementary probability measure</u> on $\mathfrak{A}$ if it satisfies the following conditions:

(m.1) $m(A \cup B) = m(A) + m(B)$ for disjoint $A, B \in \mathfrak{A}$,

(m.2) $m(S) = 1$.

An elementary probability measure m on $\mathfrak{A}$ is called <u>σ-additive</u> on $\mathfrak{A}$ if it satisfies the following condition.

(σ) If $A_1, A_2, \ldots \in \mathfrak{A}$ are disjoint and if $A = \cup_n A_n \in \mathfrak{A}$, then
$$m(A) = \sum_n m(A_n).$$

This condition is equivalent to each of the following ones.

(C) If (i) $A_1, A_2, \ldots \in \mathfrak{A}$ and (ii) $A_1 \supset A_2 \supset \ldots \to \phi$, then
$$\lim_{n\to\infty} m(A_n) = 0.$$

(C') If (i) $A_1, A_2, \ldots \in \mathfrak{A}$ (ii) $A_1 \supset A_2 \supset \ldots$ and (iii) $\inf_n m(A_n) > 0$, then $\cap_n A_n \neq \phi$.

<u>Theorem 3 (The ex~~istence~~tension theorem)</u>. Let $\mathfrak{A}$ be an algebra on S and m an elementary probability measure on $\mathfrak{A}$. Then m can be extended to a probability measure if m is σ-additive.

<u>Proof</u>. The idea of the proof is similar to the construction of the Lebesgue measure in $[0,1]$ as an extension of the notion of length. We will only sketch the proof. Define the outer m-measure m^* and the inner m-measure m_* as follows.

$$m^*(A) = \inf \{ \sum_n m(A_n) : A_n \in \mathfrak{A}, \bigcup_n A_n \supset A \},$$

$$m_*(A) = 1 - m^*(S-A), \qquad A \subset S.$$

The class $\mathfrak{M}$ of all $A \subset S$ such that $m_*(A) = m^*(A)$ is a σ-algebra on S including $\mathfrak{A}$ and the restriction

$$\mu = m^* | \mathfrak{M} \, (= m_* | \mathfrak{M})$$

is a measure on $\mathfrak{M}$. By σ-additivity of m we can prove that (i) $\mathfrak{M} \supset \mathfrak{A}$ and (ii) $\mu = m$ on $\mathfrak{A}$. Therefore μ is a probability measure on $\mathfrak{M}$ which is an extension of m.

As a corollary of this theorem we have the following in case S is a Borel space endowed with a σ-algebra $\mathcal{S}$.

<u>Theorem 3'</u>. Suppose that $\mathfrak{A}$ is an algebra on a Borel space $S = (S, \mathcal{S})$ generating the σ-algebra $\mathcal{S}$ and that m is an elementary probability measure on $\mathfrak{A}$. If m is σ-additive on $\mathfrak{A}$, then m can be extended to a unique regular probability measure on $(S, \mathcal{S})$.

<u>Proof</u>. Let μ_1 be the probability measure constructed in Theorem 3. Since $\mathfrak{M}(\mu_1) \supset \mathfrak{A}$ and so $\mathfrak{M}(\mu_1) \supset \sigma[\mathfrak{A}]$, the restriction $\mu_2 = \mu_1 | \sigma[\mathfrak{A}]$ is also a probability measure on $\sigma[\mathfrak{A}]$. The Lebesgue extension μ of μ_2 is a regular probability measure on $(S, \mathcal{S})$ which is to be constructed. If μ and μ' are such extensions, then $\mu = \mu'$ on $\sigma[\mathfrak{A}]$ by Theorem 2. Since μ and μ' are both regular, we have $\mu = \mu'$. This proves the uniqueness.

Suppose that S is a compact Hausdorff space. Let $C(S)$ be the class of all continuous real functions on S. For every regular probability measure μ we have a real-valued functional $L = L_\mu$ on $C(S)$:

$$\text{(L)} \qquad L(f) = \int_S f(x)\mu(dx),$$

which satisfies the following conditions:

$$\text{(L.1)} \qquad L(1_S) = 1 \quad (1_S \text{ is the indicator of } S),$$

$$\text{(L.2)} \qquad L(f) \geq 0 \quad \text{for} \quad f \geq 0,$$

$$\text{(L.3)} \qquad L(f+g) = L(f) + L(g).$$

Conversely we can prove the following representation theorem that is also useful to construct a regular probability measure on a compact Hausdorff space.

Theorem 4 (The representation theorem). Let S be a compact Hausdorff space and L a functional on $C(S)$. If L satisfies (L.1), (L.2) and (L.3), then there exists a unique F-regular (and so K-regular by compactness of S) probability measure μ on S for which (L) holds.

Proof. The idea of the proof is as follows. Define $\mu(K)$ for every compact K and for every open G by

$$\mu(K) = \inf\{L(f) : f \in C(S),\ f \geq 0 \text{ on } S \text{ and } f \geq 1 \text{ on } K\}$$

and

$$\mu(G) = \sup\{\mu(K) : K \text{ compact } \subset G\}.$$

Let $\mathcal{M}$ be the class of all A such that

$$\sup\{\mu(K):\ K \text{ compact} \subset A\} = \inf\{\mu(G):\ G \text{ open} \supset A\}$$

and write $\mu(A)$ for this common value. Then μ is an F-regular probability measure on S for which (L) holds.

We will present a slight generalization of the representation theorem which will be used frequently.

<u>Theorem 4'</u>. Suppose that S is a compact Hausdorff space and that E is a subclass of $C(S)$ satisfying the following conditions:

(E.1) $E \ni 1_S$,

(E.2) E is dense in $C(S)$ with respect to the maximum norm matric,

(E.3) E is closed under ~~sum~~ linear combinations with rational coefficients, i.e., if $f, g \in E$, then $\alpha f+\beta g \in E$ (α, β: rational).

If a functional $L: E \to R^1$ satisfies the conditions (L.1), (L.2) and (L.3), then there exists a unique F-regular (and so K-regular) probability measure on S for which (L) holds.

<u>Application of the extension theorem</u>.

<u>Example 1 (Probability measures on R^1)</u>. Let μ be a regular probability measure on R^1. μ is K-regular by Prohorov's theorem (2.1 Theorem 2). The function

$$F(x) = \mu(-\infty, x], \quad x \in R^1,$$

is called the distribution function of μ. It is obvious that F satisfies the following condition:

(F.1) (increasing) $F(x) \leq F(y)$ for $x \leq y$,

(F.2) (right-continuous) $F(x+) = F(x)$,

(F.3) $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$.

Theorem 5. For every function F satisfying the above three conditions we have a unique regular probability measure on R^1 whose distribution function is F.

Proof. Let $\mathcal{J}$ be the class of all intervals of the form

$$(a,b],\ (-\infty,b],\ (a,\infty) \text{ or } (-\infty,\infty)$$

and $\mathcal{O}$ the class of all disjoint finite unions of members of $\mathcal{J}$. Define a set function $m(I)$ for $I \in \mathcal{J}$ by

$$m(a,b] = F(b) - F(a),\ m(-\infty,b] = F(b),$$

$$m(a,\infty) = 1 - F(a) \text{ and } m(-\infty,\infty) = 1$$

and extend it onto $\mathcal{O}$ by additivity. Then $\mathcal{O}$ is an algebra on R^1 and m is an elementary probability measure on $\mathcal{O}$. It is obvious that $\mathcal{O}$ generates the topological σ-algebra $\mathcal{B}^1$ on R^1. For completion of the proof it is enough (by Theorem 3') to check the condition (C'). For every $I \in \mathcal{J}$ and every $\epsilon > 0$ we have a bounded $J \in \mathcal{J}$ such that

$$\bar{J} \subset I \text{ and } m(I) - m(J) < \epsilon.$$

Therefore $\mathfrak{A}$ has the same property. Suppose that $\{A_n\}$ satisfies the assumption of (C') and set $a = \inf_n m(A_n) > 0$. Take a bounded $B_n \in \mathfrak{A}$ for each A_n such that

$$\overline{B}_n \subset A_n \quad \text{and} \quad m(A_n) - m(B_n) < 2^{-n-1} a.$$

Then

$$m(A_n - \bigcap_1^n B_i) \leq \sum_1^n m(A_n - B_i) \leq \sum_1^n m(A_i - B_i) < a/2$$

and so

$$m(\bigcap_1^n B_i) > m(A_n) - \frac{a}{2} > \frac{a}{2} .$$

This implies

$$\bigcap_1^n \overline{B}_i \supset \bigcap_1^n B_i \neq \emptyset.$$

Since every $\overline{B}_i$ is compact, we have

$$\bigcap_1^\infty \overline{B}_n \neq \emptyset \quad \text{and so} \quad \bigcap_1^\infty A_n \neq \emptyset.$$

<u>Example 2</u> (Probability measures on R^n, $n < \infty$). Let μ be a regular (and so K-regular) measure on R^n. The function

$$F(x_1, x_3, \ldots, x_n) = \mu(\prod_1^n (-\infty, x_i])$$

is called the <u>distribution function</u> of μ. Let Δ_h^i denote the difference operator:

$$\Delta_h^i F(x_1, \ldots, x_n) = F(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - F(x_1, x_2, \ldots, x_n).$$

Then $\Delta_{h_1}^1, \ldots, \Delta_{h_n}^n F(x_1, \ldots, x_n)$ is the μ-measure of the rectangular set $\Pi_1^n (x_i, x_i + h_i]$ (for $h_1, h_2, \ldots, h_n > 0$). F satisfies the following conditions.

(F.1) $\Delta^1_{h_1}, \ldots, \Delta^n_{h_n} F(x_1, \ldots, x_n) \geq 0$, for $h_1, h_2, \ldots, h_n > 0$,

(F.2) (right-continuous)

$F(x_1, \ldots, x_n) \to F(a_1, \ldots, a_n)$ as $x_i \downarrow a_i$ for every i.

(F.3) $\lim\limits_{x_i \to -\infty} F(x_1, \ldots, x_n) = 0$ for each $i = 1, 2, \ldots, n$ and

$F(x_1, x_2, \ldots, x_n) \to 1$ as $x_i \to \infty$ for every i.

In the same way as in the proof of Theorem 5 we can prove the following.

Theorem 5'. Theorem 5 holds in R^n.

Example 3. (Probability measures on R^∞). Let μ be a regular probability measure on R^∞. μ is obviously K-regular by Prohorov's theorem. Let $\Pi_n: R^\infty \to R^n$ and $\Pi_{n,n+1}: R^{n+1} \to R^n$ be projection operators, i.e.,

$$\Pi_n(x_1, x_2, \ldots, x_n, x_{n+1}, \ldots) = (x_1, x_2, \ldots, x_n)$$

and

$$\Pi_{n,n+1}(x_1, x_2, \ldots, x_{n+1}) = (x_1, x_2, \ldots, x_n).$$

It is obvious that $\Pi_n = \Pi_{n,n+1} \circ \Pi_{n+1}$. Both Π_n and $\Pi_{n,n+1}$ are continuous and so Borel measurable. Let μ_n be the image measure $\Pi_n \cdot \mu$. Then we have a system of regular probability measures μ_n on R^n, $n = 1, 2, \ldots$. Obviously we have

(K) (Kolmogorov's consistency condition)

$$\mu_n = \Pi_{n,n+1} \cdot \mu_{n+1}, \quad n = 1, 2, \ldots .$$

Theorem 5" (Kolmogorov's extension theorem). For a system of regular probability measures μ_n on R^n, $n = 1, 2, \ldots$ satisfying (K), we have a unique regular probability measure μ on R^∞ such that $\mu_n = \Pi_n \cdot \mu$, $n = 1, 2, \ldots$.

Proof. Consider the class $\mathcal{O}$ of all sets A of the form $A = \Pi_n^{-1}(B)$, $B \in \mathcal{B}^n$, $n = 1, 2, \ldots$. Then $\mathcal{O}$ is an algebra on R^∞. The topological σ-algebra $\mathcal{B}^\infty$ on R^∞ is generated by $\mathcal{O}$. Define $m: \mathcal{O} \to [0,1]$ by

$$m(A) = \mu_n(B) \quad \text{for} \quad A = \Pi_n^{-1}(B).$$

Suppose that $\Pi_n^{-1}(B_1) = \Pi_{n+1}^{-1}(B_2)(B_1 \in \mathcal{B}^n,\ B_2 \in \mathcal{B}^{n+1})$. Then we have

$$\Pi_{n+1}^{-1}(B_2) = \Pi_n^{-1}(B_1) = (\Pi_{n,n+1} \circ \Pi_{n+1})^{-1}(B_1) = \Pi_{n+1}^{-1}(\Pi_{n,n+1}^{-1}(B_1)).$$

Since Π_{n+1} is a map from R^∞ onto R^{n+1}, we have

$$B_2 = \Pi_{n,n+1}^{-1}(B_1).$$

Therefore $\mu_{n+1}(B_2) = (\Pi_{n,n+1} \cdot \mu_{n+1})(B_1) = \mu_n(B_1)$. Using this several times, we have

$$\Pi_n^{-1}(B_1) = \Pi_{n+k}^{-1}(B_2)(B_1 \in \mathcal{B}^n, B_2 \in \mathcal{B}^{n+k}) \Rightarrow \mu_n(B_1) = \mu_{n+k}(B_2).$$

This implies that $m(A)$ is well-defined independently of the expression $A = \Pi_n^{-1}(B)$. It is easy to see that m is an elementary probability measure on $\mathcal{O}$. For completion of the proof it is enough to check the condition (C'). Suppose that $A_n \in \mathcal{O}$, $n = 1, 2, \ldots$ is a sequence satisfying the assumption of (C') and set $a = \inf_n m(A_n) > 0$. By inserting a number of A_n between

A_n and A_{n+1} if necessary, we can assume that ~~$A_n \in \mathcal{A}_n$, $n = 1,2,\ldots$. Write A_n as~~ $A_n = \Pi_n^{-1}(B_n)$, $B_n \in \mathcal{B}^n$. As μ_n is regular, we have a compact subset K_n of B_n such that $\mu_n(B_n - K_n) < 2^{-n-1}a$. Set $H_n = \Pi_n^{-1}(K_n)$. Then $H_n \subset A_n$ and

$$m(A_n - H_n) = \mu_n(B_n - K_n) < 2^{-n-1}a.$$

Therefore

$$m(A_n - \bigcap_1^n H_i) \leq \sum_1^n m(A_n - H_i) \leq \sum_1^n m(A_i - H_i) < a/2.$$

Since $m(A_n) > a$, we have

$$m(\bigcap_1^n H_i) > a/2 \quad \text{and so} \quad \bigcap_1^n H_i \neq \emptyset.$$

It remains only to prove that $\bigcap_1^\infty H_i \neq \emptyset$. This does not follow at once because H_i is not compact. Let us consider the space $\overline{R}^\infty$ ($\overline{R} = [-\infty, \infty]$), which is compact since $\overline{R}$ is compact. Let $\overline{\Pi}_n$ be the projection from $\overline{R}^\infty$ onto $\overline{R}^n$ and set $\overline{H}_n = \overline{\Pi}_n^{-1}(K_n)$. Since $\overline{\Pi}_n$ is continuous, $\overline{H}_n$ is closed in $\overline{R}^\infty$ and therefore compact. Since $\overline{H}_n \supset H_n$, we have

$$\bigcap_1^n \overline{H}_i \supset \bigcap_1^n H_i \neq \emptyset, \quad n = 1, 2, \ldots .$$

Since every $\overline{H}_n$ is compact, this implies that

$$\bigcap_1^\infty \overline{H}_n \neq \emptyset.$$

Take a point $x = (x_1, x_2, \ldots) \in \overline{R}^\infty$ in this intersection. Since $\overline{H}_n = \overline{\Pi}_n^{-1}(K_n)$, $n = 1, 2, \ldots,$ we have

$$(x_1, x_2, \ldots, x_n) \in K_n, \; n = 1, 2, \ldots ,$$

and so

$$x_n \in R^1, \quad n = 1, 2, \ldots, \quad \text{i.e.} \quad x \in R^\infty.$$

Then

$$x \in \Pi_n^{-1}(x_1, \ldots, x_n) \subset \Pi_n^{-1}(K_n) = H_n, \quad n = 1, 2, \ldots .$$

This completes the proof.

The Kolmogorov extension theorem can be generalized in various directions. We will present two generalizations. One is for a countable product of standard Borel spaces and the other for the projective limit of a directed countable family of probability measures.

<u>The Kolmogorov extension theorem for standard Borel spaces.</u>

Let $(S_n, \mathcal{S}_n)$, $n = 1, 2, \ldots$ be a sequence of standard Borel spaces. Then the product space $T_n = \Pi_{i=1}^{n} S_i$ with $\mathcal{T}_n = \otimes_{i=1}^{n} \mathcal{S}_i$ is also a standard Borel space for every n. Similarly for the space $T = \Pi_{i=1}^{\infty} S_i$ with $\mathcal{T} = \otimes_{i=1}^{\infty} \mathcal{S}_i$. Let us consider the following projection operators:

$$\Pi_n : T \to T_n, \; \Pi_n(x_1, x_2, \ldots) = (x_1, x_2, \ldots, x_n),$$

$$\Pi_{n,n+1} : T_{n+1} \to T_n, \; \Pi_{n,n+1}(x_1, x_2, \ldots, x_{n+1}) = (x_1, x_2, \ldots, x_n).$$

All these maps are obviously Borel measurable. Suppose that we are given a regular probability measure μ_n on $(T_n, \mathcal{T}_n)$ for every n.

<u>Theorem 6</u>. If $\mu_n = \Pi_{n,n+1} \cdot \mu_{n+1}$ for every n, then there exists a unique regular probability measure μ on $(T, \mathcal{T})$ such that $\mu_n = \Pi_n \cdot \mu$ for every n.

<u>Proof</u>. Let $\mathcal{O}\!\mathcal{L}$ be the union of the classes $\Pi_n^{-1}(\mathcal{T}_n)$, $n=1,2,\ldots$. Then $\mathcal{O}\!\mathcal{L}$ is obviously a multiplicative class on T generating $\mathcal{T}$. If we have such a measure μ on T, μ is determined on $\mathcal{O}\!\mathcal{L}$ by

$$\mu(A) = \mu_n(B) \quad \text{for} \quad A = \Pi_n^{-1}(B).$$

Therefore such a μ is unique by Theorem 2. Now we will prove the existence of μ. Since $(S_n, \mathcal{S}_n)$ is standard for every n, we can assume with no loss of generality that S_n is a Borel subset of R^1 and $\mathcal{S}_n = \mathcal{B}^1 \cap S_n$. Then T_n is a Borel subset of R^n and $\mathcal{T}_n = \mathcal{B}^n \cap T_n$. Let $\tilde{\Pi}_n: R^\infty \to R^n$ and $\tilde{\Pi}_{n,n+1}: R^{n+1} \to R^n$ be projection maps and $i_n: T_n \to R^n$ and $i: T \to R^\infty$ be identity maps. It is obvious that

$$i_n \circ \Pi_{n,n+1} = \tilde{\Pi}_{n,n+1} \circ i_{n+1} \text{ and } i_n \circ \Pi_n = \tilde{\Pi}_n \circ i;$$

in fact both maps in the first equality carry $(x_1, x_2, \ldots, x_{n+1}) \in T_{n+1}$ to $(x_1, x_2, \ldots, x_n) \in R^n$ and both maps in the second equality carry $(x_1, x_2, \ldots) \in T$ to $(x_1, x_2, \ldots, x_n) \in R^n$. Let $\tilde{\mu}_n$ denote the image measure $i_n \cdot \mu_n$. Then $\tilde{\mu}_n$ is a regular probability measure on $(R^n, \mathcal{B}^n)$. Keeping the above equalities in mind we have

$$\begin{aligned}\tilde{\Pi}_{n,n+1} \cdot \tilde{\mu}_{n+1} &= \tilde{\Pi}_{n,n+1} \cdot (i_{n+1} \cdot \mu_{n+1}) = (\tilde{\Pi}_{n,n+1} \circ i_{n+1}) \cdot \mu_{n+1} \\ &= (i_n \circ \Pi_{n,n+1}) \cdot \mu_{n+1} = i_n \cdot (\Pi_{n,n+1} \cdot \mu_{n+1}) \\ &= i_n \cdot \mu_n = \tilde{\mu}_n .\end{aligned}$$

Applying the Kolmogorov extension theorem to $\{\tilde{\mu}_n\}$, we have a regular probability measure $\tilde{\mu}$ on $(R^\infty, \mathcal{B}^\infty)$ such that

$$\tilde{\mu}_n = \tilde{\Pi}_n \cdot \tilde{\mu}.$$

It is easy to see that

$$T = \cap_n \Pi_n^{-1}(T_n) \in \mathcal{B}^\infty \quad \text{and} \quad \mathcal{T} = \mathcal{B}^\infty \cap T.$$

Since $\tilde{\mu}(\check{\Pi}_n^{-1}(T_n)) = \tilde{\mu}_n(T_n) = 1$ for every n, we have $\tilde{\mu}(T) = 1$. Therefore the restriction $\mu = \tilde{\mu}|T$ is a regular probability measure on $(T, \mathcal{T})$. It is obvious that $\tilde{\mu} = i \cdot \mu$, because $i^{-1}(A) = A \cap T$ for $A \subset R^{\infty}$. It remains only to prove

$$\mu_n = \Pi_n \cdot \mu.$$

Suppose that $B \in \mathcal{T}_n$. Then $B \subset T_n$ and so $i_n^{-1} B = B \cap T_n = B$. Therefore we have

$$\mu_n(B) = \mu_n(i_n^{-1} B) = (i_n \mu_n)(B)$$

and similarly

$$(\Pi_n \mu)(B) = (i_n \cdot (\Pi_n \cdot \mu))(B) = ((i_n \circ \Pi_n) \cdot \mu)(B).$$

But

$$i_n \mu_n = \tilde{\mu}_n = \tilde{\Pi}_n \tilde{\mu} = \tilde{\Pi}_n (i \cdot \mu) = (\tilde{\Pi}_n \circ i) \cdot \mu$$

$$= (i_n \circ \Pi_n) \cdot \mu.$$

Therefore $\mu_n(B) = (\Pi_n \mu)(B)$ for $B \in \mathcal{T}_n$. Since both measures are regular on $(T_n, \mathcal{T}_n)$, we have $\mu_n = \Pi_n \cdot \mu$.

<u>The projective limit of probability measures</u>. Let A be a directed and countably infinite set and suppose that we are given a standard probability space $(S_\alpha, \mathcal{S}_\alpha, \mu_\alpha)$ and a Borel measurable map $\varphi_{\alpha\beta}: S_\beta \to S_\alpha$ for every pair $(\alpha, \beta) \in A \times A$ with $\alpha < \beta$ satisfying the following conditions:

(D_1) $\qquad \varphi_{\alpha\beta} \circ \varphi_{\beta\gamma} = \varphi_{\alpha\gamma}$ for $\alpha < \beta < \gamma$,

(D_2) $\qquad \mu_\alpha = \varphi_{\alpha\beta} \cdot \mu_\beta$ for $\alpha < \beta$.

We want to define the <u>projective limit</u> of $(S_\alpha, \mathcal{S}_\alpha, \mu_\alpha)$, $\alpha \in A$ and denote it by

$$(S, \mathcal{S}, \mu) = \overleftarrow{\lim_\alpha} (S_\alpha, \mathcal{S}_\alpha, \mu_\alpha).$$

The projective limit $S = \overleftarrow{\lim}_\alpha S_\alpha$ is defined as usual. Let T denote the product space $\Pi_\alpha S_\alpha$, p_α the projection map from T onto S_α. The subset S of T that consists of all $x \in T$ such that

$$p_\alpha(x) = \varphi_{\alpha\beta}(p_\beta(x)) \text{ for } \alpha < \beta$$

is called the <u>projective limit</u> of S_α, $\alpha \in A$ and denoted by $\overleftarrow{\lim}_\alpha S_\alpha$.

Let $\mathcal{T}$ be the product σ-algebra $\otimes_\alpha \mathcal{S}_\alpha$ on T. The trace σ-algebra $\mathcal{S}$ of $\mathcal{T}$ on S i.e. $\mathcal{T} \cap S$ is called the <u>projective limit</u> of $\mathcal{S}_\alpha$, $\alpha \in A$ and denoted by $\overleftarrow{\lim}_\alpha \mathcal{S}_\alpha$. We call the Borel space $(S, \mathcal{S})$ the <u>projective limit</u> of $(S_\alpha, \mathcal{S}_\alpha)$, $\alpha \in A$ and denoted by $\overleftarrow{\lim}_\alpha (S_\alpha, \mathcal{S}_\alpha)$. As every Borel space $(S_\alpha, \mathcal{S}_\alpha)$ is standard, so is $(T, \mathcal{T})$ by <u>1.3</u> Theorem 4. We will prove that $(S, \mathcal{S})$ is also standard. By <u>1.3</u> Theorem 2 it is enough to show that S is a Borel subset of T, i.e. $S \in \mathcal{T}$. By the definition S is expressed as

$$S = \bigcap_{\alpha < \beta} B_{\alpha\beta} \text{ where } B_{\alpha\beta} = \{x \in T : p_\alpha(x) = (\varphi_{\alpha\beta} \circ p_\beta)(x)\}.$$

As A is countable, it is enough to show that $B_{\alpha\beta}$ is a Borel subset of T for $\alpha < \beta$. But this follows at once by 1.3 Theorem 6.

The restriction of p_α to S is denoted by φ_α. φ_α is a Borel measurable map from $(S, \mathcal{S})$ into $(S_\alpha, \mathcal{S}_\alpha)$. It is obvious that

$$\varphi_\alpha = \varphi_{\alpha\beta} \circ \varphi_\beta \quad \text{for} \quad \alpha < \beta.$$

The regular probability measure μ on S introduced in the following theorem is called the projective limit of μ_α, $\alpha \in A$, $\overleftarrow{\lim}_\alpha \mu_\alpha$ in notation.

Theorem 7. There exists a unique regular probability measure μ on $(S,\mathcal{S})$ such that

$$\mu_\alpha = \varphi_\alpha \cdot \mu, \quad \alpha \in A.$$

Proof. Number the elements of A as $\{\alpha_1, \alpha_2, \ldots\}$ and set

$$T_n = \Pi_{i=1}^n S_{\alpha_i} \quad \text{and} \quad \mathcal{T}_n = \otimes_{i=1}^n \mathcal{S}_{\alpha_i}, \quad i = 1, 2, \ldots .$$

We will consider the following Borel measurable maps:

$$\Pi_n \colon T \to T_n, \quad \Pi_n x = (p_{\alpha_1}(x), \ldots, p_{\alpha_n}(x)).$$

$$\Pi_{n,n+1} \colon T_{n+1} \to T_n, \quad \Pi_{n,n+1}(x_1, x_2, \ldots, x_{n+1}) = (x_1, x_2, \ldots, x_n).$$

$$p_{\alpha,n} \colon T_n \to S_\alpha, \quad p_{\alpha,n}(x_1, x_2, \ldots, x_n) = x_i \quad \text{if} \quad \alpha = \alpha_i \quad (i \leq n).$$

$$f_{n\alpha} : S_\alpha \to T_n, \quad f_{n\alpha}(z) = (\varphi_{\alpha_1\alpha}(z), \varphi_{\alpha_2\alpha}(z), \ldots, \varphi_{\alpha_n\alpha}(z))$$

$$\text{if} \quad \alpha > \alpha_1, \alpha_2, \ldots, \alpha_n.$$

$$f_n \colon S \to T_n, \quad f_n(x) = (p_{\alpha_1}(x), \ldots, p_{\alpha_n}(x)) = (\varphi_{\alpha_1}(x), \ldots, \varphi_{\alpha_n}(x)).$$

The following obvious relations will be frequently used:

$$f_n = \Pi_n | S,$$

$$p_\alpha = p_{\alpha n} \circ \Pi_n \quad \text{if} \quad \alpha = \alpha_i \ (i \leq n),$$

$$\varphi_{\alpha\beta} = p_{\alpha n} \circ f_{n\beta} \quad \text{if} \quad \alpha = \alpha_i \ (i \leq n) \quad \text{and} \quad \beta > \alpha_1, \alpha_2, \ldots, \alpha_n,$$

$$f_{n\gamma} = f_{n\alpha} \circ \varphi_{\alpha\gamma} \quad \text{if} \quad \alpha_1, \alpha_2, \ldots, \alpha_n < \alpha < \gamma,$$

$$f_{n\alpha} = \Pi_{n,n+1} \circ f_{n+1,\alpha} \quad \text{if} \quad \alpha > \alpha_1, \alpha_2, \ldots, \alpha_n, \alpha_{n+1},$$

$$f_n = f_{n\alpha} \circ \varphi_\alpha \quad \text{if} \quad \alpha_1, \alpha_2, \ldots, \alpha_n < \alpha.$$

Suppose that μ is a regular probability measure on $(S, \mathcal{S})$ such that $\mu_\alpha = \varphi_\alpha \cdot \mu$ for $\alpha \in A$. Then if $\alpha_1, \alpha_2, \ldots, \alpha_n < \alpha$, we have

$$f_n \cdot \mu = (f_{n\alpha} \circ \varphi_\alpha) \cdot \mu = f_{n\alpha}(\varphi_\alpha \cdot \mu) = f_{n\alpha} \cdot \mu_\alpha .$$

Therefore for $E \in \mathcal{T}_n$ we have

$$\mu(f_n^{-1}(E)) = (f_n \cdot \mu)(E) = (f_{n\alpha} \cdot \mu_\alpha)(E),$$

namely $\mu(f_n^{-1}(E))$, $E \in \mathcal{T}_n$, $n = 1, 2, \ldots$ are determined by μ_α, $\alpha \in A$. Since $f_n^{-1}(E)$, $E \in \mathcal{T}_n$, $n = 1, 2, \ldots$ form a multiplicative class generating $\mathcal{S}$, the measure μ is determined by μ_α, $\alpha \in A$ by Theroem 2. This proves the uniqueness part of our theorem.

The above observation also gives a clue of the existence proof. Define a regular probability measure ν_n on $(T_n, \mathcal{T}_n)$ by

$$\nu_n = f_{n\alpha} \cdot \mu_\alpha ,$$

where α is any element of $\alpha > \alpha_1, \alpha_2, \ldots, \alpha_n$. The measure ν_n is well-defined independently of the choice of α by the following fact:

$$f_{n\alpha} \cdot \mu_\alpha = f_{n\beta}\, \mu_\beta \quad \text{for} \quad \alpha_1, \alpha_2, \ldots, \alpha_n < \alpha, \beta .$$

Take $\gamma > \alpha, \beta$. Then

$$f_{n\alpha} \cdot \mu_\alpha = f_{n\alpha}(\varphi_{\alpha\gamma} \cdot \mu_\gamma) = (f_{n\alpha} \circ \varphi_{\alpha\gamma}) \cdot \mu_\gamma = f_{n\gamma} \cdot \mu_\gamma .$$

Similarly

$$f_{n\beta} \cdot \mu_\beta = f_{n\gamma} \cdot \mu_\gamma \quad \text{and so} \quad f_{n\alpha} \cdot \mu_\alpha = f_{n\beta} \cdot \mu_\beta .$$

We will use Theorem 6 to obtain a regular probability measure ν on $(T, \mathfrak{J})$ such that

$$\nu_n = \Pi_n \cdot \nu, \quad n = 1, 2, \ldots .$$

For this purpose it is enough to observe

$$\nu_n = f_{n\alpha} \cdot \mu_\alpha = (\Pi_{n,n+1} \circ f_{n+1,\alpha}) \cdot \mu_\alpha$$

$$= \Pi_{n,n+1}(f_{n+1,\alpha} \cdot \mu_\alpha) = \Pi_{n,n+1} \cdot \nu_{n+1}$$

where $\alpha > \alpha_1, \alpha_2, \ldots, \alpha_{n+1}$.

First we will prove that

$$\mu_\alpha = p_\alpha \cdot \nu, \quad \alpha \in A.$$

Take n such that $\alpha = \alpha_n$ and then take $\beta > \alpha$. Then

$$p_\alpha \nu = (p_{\alpha n} \circ \Pi_n) \cdot \nu = p_{\alpha n}(\Pi_n \cdot \nu) = p_{\alpha n} \cdot \nu_n$$

$$= p_{\alpha n}(f_{n\beta} \cdot \mu_\beta) = (p_{\alpha n} \circ f_{n\beta}) \cdot \mu_\beta = \varphi_{\alpha\beta} \cdot \mu_\beta = \mu_\alpha .$$

Second we will prove that

$$\nu(S) = 1.$$

Since $S = \bigcap_{\alpha < \beta} B_{\alpha\beta}$ where

$$B_{\alpha\beta} = \{x \in T: p_\alpha(x) = (\varphi_{\alpha\beta} \circ p_\beta)(x)\},$$

it is enough to show

$$\nu(B_{\alpha\beta}) = 1 \quad \text{for} \quad \alpha < \beta.$$

Take n such that $\alpha, \beta \subset \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and then take $\gamma > \alpha_1, \alpha_2, \ldots, \alpha_n$. Then

$$\nu(B_{\alpha\beta}) = \nu\{x \in T: p_{\alpha n}(\Pi_n x) = (\varphi_{\alpha\beta} \circ p_{\beta n})(\Pi_n(x))\}$$

$$= \nu_n\{y \in T_n: p_{\alpha n}(y) = (\varphi_{\alpha\beta} \circ p_{\beta n})(y)\} \quad \text{by} \quad \nu_n = \Pi_n \cdot \nu$$

$$= \mu_\gamma\{z \in S_\gamma: p_{\alpha n}\, f_{n\gamma}(z) = (\varphi_{\alpha\beta} \circ p_{\beta n})(f_{n\gamma}(z))\} \quad \text{by} \quad \nu_n = f_{n\gamma} \cdot \mu_\gamma$$

$$= \mu_\gamma\{z \in S_\gamma: \varphi_{\alpha\gamma}(z) = (\varphi_{\alpha\beta} \circ \varphi_{\beta\gamma})(z)\}$$

$$= \mu_\gamma\{z \in S_\gamma: \varphi_{\alpha\gamma}(z) = \varphi_{\alpha\gamma}(z)\} = 1.$$

Define μ to be the restriction of ν to S. Then μ is a regular probability measure on $(S, \mathcal{S})$. Let $i: S \to T$ be the identity operator. Then we have

$$i \cdot \mu = \nu \quad \text{and} \quad \varphi_\alpha = p_\alpha \circ i,$$

and so

$$\varphi_\alpha \cdot \mu = (p_\alpha \circ i)\mu = p_\alpha(i\mu) = p_\alpha \nu = \mu_\alpha.$$

This shows that μ is the measure we wanted to construct.

Remark. Theorem 6 is obviously a generalization of the Kolmogorov extension theorem. Now we will explain that Theorem 6 is an immediate consequence of Theorem 7. The spaces $(T_n, \mathcal{T}_n, \mu_n)$, $n = 1,2,\ldots$ in Theorem 6 satisfy the conditions in Theorem 7 with respect to

$$A = \{1,2,3,\ldots\} \text{ with the natural ordering}$$

and

$$\varphi_{ij} = \Pi_{i,i+1} \circ \Pi_{i+1,i+2} \circ \cdots \circ \Pi_{j-1,j} \quad \text{for } i < j.$$

Therefore we have the projective limit

$$(\tilde{S}, \tilde{\mathcal{B}}, \tilde{\mu}) = \overleftarrow{\lim}_n (T_n, \mathcal{T}_n, \mu_n)$$

by Theorem 7. Every $x \in \tilde{S}$ is written uniquely as

$$x = (x_1, (x_1, x_2), (x_1, x_2, x_3), \ldots)$$

with $x_i \in S_i$, $i = 1,2,\ldots$ and conversely every x of this form is in $\tilde{S}$. It is easy to see that the map $\varphi: \tilde{S} \to T = \Pi_n S_n$ that carries the above x to $(x_1, x_2, \ldots)$ gives a Borel isomorphic map from $(\tilde{S}, \tilde{\mathcal{B}})$ onto $(T, \mathcal{T}) = (\Pi_n S_n, \otimes_n \mathcal{B}_n)$. The image measure $\mu = \varphi \cdot \tilde{\mu}$ is the measure we wanted to construct in Theorem 6.

Application of the representation theorem

Let S_λ, $\lambda \in \Lambda$ be a family of compact Hausdorff topological spaces. Then the product space $S_M = \Pi_{\lambda \in M} S_\lambda$ with the product topology is also compact Hausdorff for every subset M of Λ. For $M_1 \subset M_2$ we write $\Pi_{M_1 M_2}$ for the natural projection from S_{M_2} onto S_{M_1}. Since $\Pi_{M_1 M_2}$ is continuous, the image $\Pi_{M_1 M_2}(K)$ is compact if K is compact. Using this fact we can easily prove that if μ_2 is a F-regular (and so K-regular) probability measure on S_{M_2}, then the image measure $\mu_1 = \Pi_{M_1 M_2} \cdot \mu_2$ is also K-regular. Analogously to the Kolmogorov extension theorem we have the following.

<u>Theorem 8</u>. Suppose we are given a F-regular (and so K-regular) probability measure μ_M on S_M for every finite subset M of Λ. If the system $\{\mu_M\}$ satisfies the Kolmogorov consistency condition:

$$M_1 \subset M_2 \Rightarrow \mu_{M_1} = \Pi_{M_1 M_2} \cdot \mu_{M_2} ,$$

then we have a unique K-regular probability measure μ_Λ on S_Λ such that

$$\mu_M = \Pi_{M\Lambda} \cdot \mu_\Lambda \qquad \text{for every finite set } M \subset \Lambda .$$

<u>Proof</u>. A real function f on S_Λ is called a <u>tame function</u> if we have a finite subset M of Λ and a function g on S_M such that

(T) $\qquad f = g \circ \Pi_{M\Lambda}$.

f is measurable $\mathcal{B}(S_\Lambda)$ if and only if g is measurable $\mathcal{B}(S_M)$. Let $\underline{\underline{T}}$ be the space of all bounded tame functions measurable $\mathcal{B}(S_\Lambda)$. It is obvious that $\underline{\underline{T}}$ is a real vector space. For $f \in \underline{\underline{T}}$ we set

$$I(f) = \int_{S_M} g(y)\mu_M(dy).$$

$I(f)$ is well defined independently of the expression (T) by virtue of the consistency condition. It is easy to see that I is a bounded positive linear functional on $\underline{\underline{T}}$ with $I(1) = 1$.

Every $f \in C(S_\Lambda)$ is expressed as a uniformly convergent limit of a sequence $f_n \in \underline{\underline{T}}$, $n = 1,2,\ldots$. If $f \geq 0$, we can take $f_n \geq 0$, $n = 1,2,\ldots$. We will prove this fact. For $x \in S$ we can find a neighborhood $U_n(x)$ such that

$$f(U_n(x)) \subset (f(x)-1/n,\ f(x)+1/n).$$

By the definition of product topology we can assume that the indicator of $U_n(x)$ belongs to $\underline{\underline{T}}$. By the covering theorem we have

$$S_\Lambda \subset \bigcup_{i=1}^{k(n)} U_n(x_i) \qquad (k(n) < \infty).$$

Let

$$f_n(x) = f(x_i) \quad \text{on} \quad U_n(x_1)^c \cap \ldots \cap U_n(x_{i-1})^c \cap U_n(x_i),$$

$$(i = 1,2,\ldots,k(n)).$$

Then $|f_n(x)-f(x)| < 1/n$, $x \in S_\Lambda$. Let e_{ni} denote the indicator of $U_n(x_i)$. Then

$$f_n(x) = \sum_{i=1}^{k(n)} f(x_i)(1-e_{ni}(x))\ldots(1-e_{n,i-1}(x))e_{ni}(x).$$

Therefore $f_n \in \underline{\underline{T}}$, and $f_n \geq 0$ for $f \geq 0$.

Now we will define $L(f)$ for $f \in C(S_\Lambda)$. Take a sequence $f_n \in \underline{\underline{T}}$ convergent uniformly to f. Then

$$|I(f_n)-I(f_m)| \leq \| f_n-f_m \|_\infty , \quad \| f \|_\infty = \sup_x |f(x)|$$

$$\leq \| f_n-f \|_\infty + \| f_m-f \|_\infty \to 0$$

as $n, m \to \infty$. Set

$$L(f) = \lim_n I(f_n).$$

$L(f)$ is well defined independently of the choice of $\{f_n\}$. As I is bounded and linear, so is L. If $f \geq 0$, we can take $f_n \geq 0$, so that $L(f) \geq 0$. $L(1) = 1$ is obvious.

By the representation theorem we have a K-regular probability measure μ_Λ on S_Λ such that

$$L(f) = \int_{S_\Lambda} f(x)\mu_\Lambda(dx).$$

It is easy to prove that $\mu_M = \Pi_{M\Lambda}\, \mu_\Lambda$ for every finite subset M of Λ. It is also easy to prove the uniqueness.

2.3. The Mathematical Set-up of Probability Theory

To study a random phenomenon we take a probability space (Ω,P) where Ω is the space of all possible outcomes in observing the phenomenon and $P(A)$ indicates the probability that the observed outcome drops in A. Ω is called the sample space of the random phenomenon, a point in Ω a sample point and P the probability law governing the phenomenon. Take a generic point ω in Ω. A condition $\alpha = \alpha(\omega)$ depending on ω is called an event if the set $\{\omega:\alpha(\omega)\}$ is P-measurable, i.e. if this set belongs to $\mathcal{M}(P)$. The P-measure of this set is called the probability (of occurrence) of α and denoted by $P(\alpha)$. Let $S = (S,\mathcal{S})$ be a Borel space. A function $X(\omega)$ with values in S is called an $(S,\mathcal{S})$-valued random variable if it is P-measurable i.e., if it is measurable $\mathcal{M}(P)/\mathcal{S}$. The Borel space $(S,\mathcal{S})$ is called the range space or sample space of the random variable X. According as the range space is R^1, R^n, R^∞ or a function space, the random variable is called a real random variable, an n-dimensional random vector, a random sequence or a random function. Starting with these basic notions we can introduce all other notions in terms of measure theory.

In this book we assume the following.

(A.1) The sample space is a standard Borel space $(\Omega,\mathcal{F})$ and the probability law P is a regular probability measure on $(\Omega,\mathcal{F})$.

(A.2) The range space of every random variable in consideration is a standard Borel space.

As we have seen in Chapter 1, practically every space that may appear as a sample space is a standard Borel space, so that these restrictions are harmless. On the other hand these restrictions will enable us to formulate many facts in probability theory more naturally.

Suppose that we want to study the <u>coin tossing game</u> in which we flip a coin infinitely many times and observe the sides coming out. Every sample point is an infinite sequence with terms 0 or 1: 0 stands for tail and 1 for head, so that the sequence $(0,1,1,0,1,\ldots)$ means that tail comes out first, head second, head third, tail fourth, head fifth and so on. The sample space Ω is $\{0,1\}^{\infty}$ and is a compact metrizable space with the usual product topology. Therefore Ω is a standard Borel space with the topological σ-algebra $\mathcal{F} = \mathcal{B}(\Omega)$. Let $E(i_1,i_2,\ldots,i_n)$ denote the subset of Ω defined by

$$E(i_1,i_2,\ldots,i_n) = \{\omega = (\omega_1,\omega_2,\ldots): \omega_k = i_k;\ k = 1,2,\ldots,n\}.$$

The class $\mathfrak{A}$ of all finite unions of such sets is obviously an algebra generating $\mathcal{F}$. It is natural to define

$$P(E(i_1,i_2,\ldots,i_n)) = 2^{-n}.$$

This can be extended to an elementary probability measure on $\mathfrak{A}$ by additivity. We will further extend this to a regular probability measure on $(\Omega,\mathcal{F})$ using

2.2 Theorem 3'. Suppose that (i) $A_n \in \mathfrak{A}$, $n = 1,2,\ldots,$ (ii) $A_1 \supset A_2 \supset \ldots$ and (iii) $\inf P(A_n) > 0$. As $A_n \in \mathfrak{A}$, A_n is closed in Ω and so compact. Since we have

$$\bigcap_1^n A_i = A_n \neq \emptyset$$

by (ii) and (iii), we get $\bigcap_1^\infty A_n \neq \emptyset$, as is desired. We can discuss the coin-tossing game on the probability space $(\Omega, \mathcal{F}, P)$. Let $X_n(\omega) = \omega_n$ for $\omega = (\omega_1, \omega_2, \ldots)$. Then $X_n(\omega)$ is a random variable taking 0 or 1 according as the n-th outcome is tail or head. $S_n(\omega) = \Sigma_1^n X_i(\omega)$ is a random variable with values in $\{0,1,2,\ldots,n\}$ which indicates the number of heads coming out for the first n throws.

Let us go back to general discussions. If $P(\alpha) = 1$, we say that $\alpha(\omega)$ occurs a.s. (= almost surely). Let $\alpha_n(\omega)$, $n = 1,2,\ldots$ be a sequence of events and A_n denote the set $\{\omega: \alpha_n(\omega)\}$ for $n = 1,2,\ldots$. The following equalities are often useful for computation of probabilities.

$$\{\omega: \alpha_n(\omega) \text{ occurs for some } n\} = \bigcup_n A_n$$

$$\{\omega: \alpha_n(\omega) \text{ occurs for every } n\} = \bigcap_n A_n$$

$$\begin{aligned} &\{\omega: \alpha_n(\omega) \text{ occurs for an infinite number of } n\text{'s}\} \\ &= \{\omega: \alpha_n(\omega) \text{ i.o.}\} \quad (\text{i.o.} = \text{infinitely often}) \\ &= \bigcap_n \bigcup_{k \geq n} A_k = \limsup_{n \to \infty} A_n \end{aligned}$$

$$\begin{aligned} &\{\omega: \alpha_n(\omega) \text{ occurs except for a finite number of } n\text{'s}\} \\ &= \bigcup_n \bigcap_{k \geq n} A_k = \liminf_{n \to \infty} A_n. \end{aligned}$$

<u>Borel-Cantelli's lemma</u>. If $\Sigma_n P(A_n) < \infty$, then $P(\limsup_n A_n) = 0$. In other words, if $\Sigma_n P(\alpha_n) < \infty$, then α_n occurs f.o. almost surely. (f.o. = finitely often.)

<u>Proof</u>. $P(\limsup_n A_n) \leq P(\cup_{k\geq n} A_k) \leq \Sigma_{k\geq n} P(A_k)$ for every n. The right hand side tends to 0 as $n \to \infty$ by our assumption.

Let $X(\omega)$ be an $(S, \mathcal{S})$-valued <u>random variable</u> on $(\Omega, \mathcal{F}, P)$. If and only if $\{\omega: X(\omega) \in B\} \in \mathcal{M}(P)$ $(B \subset S)$, we can define the probability of the event $X(\omega) \in B$. Denote the probability of this event by $P_X(B)$. In other words we define P_X by

$$P_X(B) = P(X^{-1}(B)) \quad \text{for } B \subset S \quad \text{such that } X^{-1}(B) \in \mathcal{M}(P)$$

i.e. P_X is the image measure of P by the map X. The probability measure P_X on S is called the probability law of the map X. By <u>2.1</u> Theorem 3 we have the following theorem that does not always hold without our assumptions (A.1) and (A.2).

<u>Theorem 1</u>. The probability law of a random variable is a regular probability measure on the range space of the variable.

Suppose that X and Y are $(S, \mathcal{S})$-valued random variables. Y is called equivalent to X if $Y(\omega) = X(\omega)$ a.s. We will slightly extend this notion to define equivalence of random variables with different range spaces. Let X and Y be random variables with values in $(S, \mathcal{S})$ and $(T, \mathcal{T})$ respectively. Y is called <u>equivalent</u> to X or a <u>version</u> of X if $Y(\omega) = X(\omega)$ a.s. It is obvious that $P_X(S \cap T) = P_Y(S \cap T) = 1$ in this case. Equivalence of random variables satisfies the usual conditions of equivalence relation.

Let μ be a probability measure on an abstract space S and S_1 a subset of S with $\mu(S_1) = 1$. The restriction of μ to $\mathcal{M}(\mu) \cap S_1$, denoted by $\mu|S_1$, is also a probability measure. μ is completely determined by $\mu|S_1$ as follows:

$$\mu(A) = (\mu|S_1)(A \cap S_1), \quad A \in \mathcal{M}(\mu).$$

Let μ and ν be probability measures on S and T respectively. If we have $U \subset S \cap T$ with $\mu(U) = \nu(U) = 1$ such that $\mu|U = \nu|U$, then we say that μ and ν are <u>equivalent</u> to each other. Equivalence of probability measures also satisfies the usual condition of equivalence relation.

<u>Theorem 2</u>. If X and Y are equivalent, then their probability laws are equivalent.

<u>Proof</u>. Set $\Omega_1 = \{\omega: X(\omega)=Y(\omega)\}$. Then $X(\Omega_1) = Y(\Omega_1)$. Denote the common image by U. Then $X^{-1}(U) \supset \Omega_1$. Since $P(\Omega_1) = 1$ by the assumption, we have $P(X^{-1}(U)) = 1$ i.e. $P_X(U) = 1$. Similarly $P_Y(U) = 1$. It is now easy to see $P_X|U = P_Y|U$ by noticing that $X^{-1}(A) \cap \Omega_1 = Y^{-1}(A) \cap \Omega_1$.

Let $(S, \mathcal{B})$ be a standard Borel space. If $X: \Omega \to S$ is Borel measurable (i.e. measurable $\mathcal{F}/\mathcal{B}$) and if $X(\Omega) = S$, then X is called a <u>standard random variable</u>.

<u>Theorem 3</u>. For every $(S, \mathcal{B})$-valued random variable $X(\omega)$ we have a standard version with values in a set $T \in \mathcal{B}$ i.e. a standard $(T, \mathcal{B} \cap T)$-valued random variable equivalent to X.

Proof. We can assume with no loss of generality that S is a Borel subset of $(0,1]$ and $\mathcal{S} = \mathcal{B}^1 \cap S$. Write E_{nk} for $((k-1)/n, k/n] \cap S$ and A_{nk} for $X^{-1}(E_{nk})$. Take a set $B_{nk} \in \mathcal{F}$ such that $B_{nk} \subset A_{nk}$ and $P(A_{nk} - B_{nk}) = 0$. Set $B = \cap_n \cup_k B_{nk}$. Then $B \in \mathcal{F}$ and $P(B) = 1$ because $P(\cup_k B_{nk}) = 1$ for every n. Take a point $a \in S$ and fix it for the moment. Define $X_n(\omega)$ by

$$X_n(\omega) = \begin{cases} \text{any fixed point in } E_{nk} & \text{for } \omega \in B_{nk} \cap B \ (k = 1,2,\dots n) \\ a & \text{for } \omega \in \Omega - B; \end{cases}$$

notice that if $\omega \in B_{nk} \cap B$, then $B_{nk} \cap B \neq \emptyset$ and so $E_{nk} \neq \emptyset$. $X_n: \Omega \to S$ is obviously measurable $\mathcal{F}/\mathcal{S}$. Since $|X_n(\omega) - X(\omega)| \leq 1/n$ on B and $X_n(\omega) = a$ on $\Omega - B$, $Y(\omega) = \lim_n X_n(\omega)$ exists for every ω and we have

$$Y(\omega) = X(\omega) \text{ on } B \text{ and } = a \text{ on } \Omega - B.$$

As $X_n: \Omega \to S$ is measurable $\mathcal{F}/\mathcal{S}$ for every n, so is Y. Since $P(B) = 1$, we have $P(X = Y) = 1$. For completion of the proof it is enough to modify Y to obtain a standard version Z of X. Since $Y^{-1}(Y(\Omega)) = \Omega$, $P_Y(Y(\Omega)) = 1$ and we have a set $T \in \mathcal{S}$ such that $P_Y(T) = 1$. Define $Z: \Omega \to T$ by

$$Z(\omega) = \begin{cases} Y(\omega) & \text{on } Y^{-1}(T) \\ \text{any fixed point in } T & \text{on } \Omega - Y^{-1}(T). \end{cases}$$

Since Y is measurable $\mathcal{F}/\mathcal{S}$, we have $Y^{-1}(T) \in \mathcal{F}$. It is easy to check that Z is measurable $\mathcal{F}/\mathcal{S} \cap T$. Since

$P(Y^{-1}(T)) = P_Y(T) = 1$, we have $P(Y = Z) = 1$ and so $P(X = Z) = 1$. It is obvious that $Z(\Omega) = T$. Therefore Z is a standard version of X with values in $(T, \mathcal{S} \cap T)$.

Let X be a real or complex random variable. The integral of $X(\omega)$ over $A \in \mathfrak{M}(P)$, if exists, is denoted by $E(X,A)$. $E(X,\Omega)$ is denoted simply by $E(X)$ and is called the expectation or mean value of X. The properties of $E(X,A)$ and $E(X)$ can be derived from those of integrals.

The space of all complex random variables with

$$\| X\|_p = E(|X|^p)^{1/p} < \infty \qquad (1 \leq p < \infty)$$

is called the L^p-space over $(\Omega, \mathcal{F}, P)$, $L^p(\Omega, \mathcal{F}, P)$ or $L^p(\Omega)$ in notation, where two equivalent random variables are identified. $L^p(\Omega)$ is a Banach space. The class of all real random variables in $L^p(\Omega)$ is denoted by $L^p_R(\Omega, \mathcal{F}, P)$ or $L^p_R(\Omega)$, It is a real Banach space. In particular $L^2(\Omega)$ is a Hilbert space with the inner product:

$$(X,Y) = E(X \cdot \overline{Y}).$$

$L^2_R(\Omega)$ is a real Hilbert space similarly.

The class $\mathfrak{M}(P)$ is a complete metric space with the measure metric:

$$\rho(A,B) = P(A \ominus B)$$

where two sets are identified if they are equivalent i.e. if they differ only by a null set.

Since $(\Omega, \mathfrak{F})$ is Borel isomorphic with a Borel subset of R^1 and since P is regular, we can easily prove the following.

Theorem 4. $\mathfrak{M}(P)$, $L^p_R(\Omega)$ and $L^p(\Omega)(1 \leq p < \infty)$ are separable.

Let $X(\omega)$ be an $(S, \mathcal{S})$-valued random variable. An event $\alpha(\omega)$ is said to be determined by the value of $X(\omega)$, if $X(\omega_1) = X(\omega_2)$ implies $\alpha(\omega_1) \Longleftrightarrow \alpha(\omega_2)$. Similarly a $(T, \mathfrak{T})$-valued random variable $Y(\omega)$ is said to be determined by the value of $X(\omega)$, if $X(\omega_1) = X(\omega_2)$ implies $Y(\omega_1) = Y(\omega_2)$.

Let $\beta(s)$ be an event on the probability space $(S, \mathcal{S}, P_X)$. Then $\beta(X(\omega))$ is an event on $(\Omega, \mathfrak{F}, P)$ determined by the value of $X(\omega)$. Let $g(s)$ be a $(T, \mathfrak{T})$-valued random variable on $(S, \mathcal{S}, P_X)$. Then $g(X(\omega))$ is a $(T, \mathfrak{T})$-valued random variable on $(\Omega, \mathfrak{F}, P)$ (determined by the value of $X(\omega)$). We will prove the converse of this fact.

Theorem 5. (i) If $\alpha(\omega)$ is determined by the value of $X(\omega)$, then we have an event $\beta(s)$ on $(S, \mathcal{S}, P_X)$ such that $\alpha(\omega) = \beta(X(\omega))$.
(ii) If $Y(\omega)$ is a $(T, \mathfrak{T})$-valued random variable determined by the value of $X(\omega)$, then we have a $(T, \mathfrak{T})$-valued random variable $g(s)$ on $(S, \mathcal{S}, P_X)$ such that $Y(\omega) = g(X(\omega))$.

Proof. For an event $\alpha(\omega)$ we consider a random variable $1_\alpha(\omega)$ with values in $\{0,1\}$ as follows:

$1_\alpha(\omega) = 1$ or 0 according as $\alpha(\omega)$ occurs or not. Then $1_\alpha(\omega) = 1_\beta(\omega)$ if and only if $\alpha(\omega) \Longleftrightarrow \beta(\omega)$. Keeping this in mind we can derive (i) from (ii) at once. To prove (ii), define $g: S \to T$ by

$$g(s) = \begin{cases} Y(\omega), & \text{if } s = X(\omega) \\ \text{any fixed point in } \mathfrak{T} & \text{if } s \notin X(\Omega) \end{cases}$$

Since $Y(\omega)$ is determined by the value of $X(\omega)$, g is well-defined. It is obvious that $Y(\omega) = g(X(\omega))$. For every $B \in \mathfrak{T}$ we have

$$X^{-1}(g^{-1}(B)) = (g \circ X)^{-1}(B) = Y^{-1}(B) \in \mathfrak{M}(P).$$

Therefore $g^{-1}(B) \in \mathfrak{M}(P_X)$. This shows that g is measurable $\mathfrak{M}(P_X)/\mathfrak{T}$. The rest of the proof is trivial.

The g in the above theorem is uniquely determined only on $X(\Omega)$. Similarly for the β. Therefore we have the following.

<u>Theorem 5'</u>. If X is a standard random variable, then the β and the g in Theorem 5 are unique.

By the transformation formula on integrals we will get the following theorem immediately.

<u>Theorem 6</u>. In the situation of Theorem 5 (or 5') we have

$$P(\alpha(\omega)) = P_X(\beta(s)), \quad P(Y(\omega) \in B) = P_X(g(s) \in B)$$

and

$$E(Y) = E_X(g)\left(= \int_S g(s)\, P_X(ds)\right) \quad \text{in case } Y \text{ is complex.}$$

The β and the g introduced above are called <u>representations</u> of α and Y respectively. In view of the above theorems we can discuss all events and all random variables determined by the value of $X(\omega)$ on the sample space $(S, \mathcal{S})$ of X endowed with the probability law P_X by considering their representations.

2.4. Conditional probability measures

Let $(\Omega, \mathcal{F}, P)$ be a standard probability space. For any given $E \in \mathcal{M}(P)$ with $P(E) > 0$ the conditional probability measure P^E under E is defined by

$$(1) \qquad P^E(A) = P(E \cap A)/P(E)$$

where A moves over all sets A with $E \cap A \in \mathcal{M}(P)$. It is easy to check that P^E is a regular probability measure on $(\Omega, \mathcal{F})$ with $\mathcal{M}(P^E) \supset \mathcal{M}(P)$ concentrated on E. Since every event α is represented by the set $E = \{\omega : \alpha(\omega)\}$, the conditional probability measure P^α under α can be defined by the above formula.

Let $X(\omega)$ be an $(S, \mathcal{S})$-valued random variable. We will define the conditional probability measures $P^{X=s}$, $s \in S$. If $P(X = s) > 0$, we have

$$P^{X=s}(A) = P(\{\omega : X(\omega) = s\} \cap A)/P(X = s).$$

This does not work in general, because it may happen that $P(X = s) = 0$ for some $s \in A$ and often even for every $s \in S$. We will define $P^{X=s}$, $s \in S$ by the following three conditions.

(CP.1) $P^{X=s}$ is a regular probability measure on $(\Omega, \mathcal{F})$ for every $s \in S$.

(CP.2) $P^{X=s}(A)$ is Borel measurable in $s \in (S, \mathcal{S})$ for every $A \in \mathcal{F}$.

(CP.3) $\int_B P^{X=s}(A)\, P_X(ds) = P(A \cap (X \in B))$, $A \in \mathcal{F}$, $B \in \mathcal{S}$.

cap

(CP.1) is an obvious requirement. (CP.2) is imposed to make the integral in (CP.3) meaningful. (CP.3) is the integral formulation of the symbolic equation:

$$P^{X=s}(A) = \frac{P(A \cap (X \in ds))}{P(X \in ds)} .$$

To show that $P^{X=s}$ is well-defined, we have to prove the following two statements.

(a) There exists such a family $\{P^{X=s}\}_s$.

(b) If we have two such families, say $\{P_1^{X=s}\}_s$ and $\{P_2^{X=s}\}_s$, then

$$P_1^{X=s} = P_2^{X=s} \quad \text{a.e. } (P_X) \text{ on } S.$$

<u>Proof of (a)</u>. Fix $A \in \mathcal{F}$ for the moment and consider a set function

$$M(B) = P(A \cap X^{-1}(B)), \qquad B \in \mathcal{S}.$$

Since

$$M(B) \leq P(X^{-1}(B)) = P_X(B), \qquad B \in \mathcal{S},$$

M is absolutely continuous with respect to P_X. By the Radon-Nikodym theorem we have a Borel measurable function $\tilde{\mu}_s = \tilde{\mu}_s(A)$ of s such that $0 \leq \tilde{\mu}_s(A) \leq 1$ and that

$$(2) \qquad \int_B \tilde{\mu}_s(A) P_X(ds) = P(A \cap X^{-1}(B)), \quad A \in \mathcal{F},\ B \in \mathcal{S}.$$

Since $(\Omega, \mathcal{F})$ is a standard Borel space, we can assume with no loss of generality that Ω is a Borel subset of R^1 and $\mathcal{F} = B^1 \cap \Omega$. We assume that the symbols r and r' below always denote rationals. Set

$$\tilde{F}_s(r) = \tilde{\mu}_s((-\infty, r] \cap \Omega).$$

$\tilde{F}_s(r)$ is obviously Borel measurable in $s \in S$ for each r. Then we have

$$\int_B \tilde{F}_s(r) P_X(ds) = P((-\infty, r] \cap X^{-1}(B)) \tag{2'}$$

by setting $A = (-\infty, r] \cap \Omega$ in (2) and noticing $X^{-1}(B) \subset \Omega$. Now consider the following Borel subsets of S:

$$S(r, r') = \{s \in S : \tilde{F}_s(r) \leq \tilde{F}_s(r')\}, \quad r < r'.$$

Since $r < r'$, we have

$$P((-\infty, r] \cap X^{-1}(B)) \leq P((-\infty, r'] \cap X^{-1}(B))$$

i.e.

$$\int_B \tilde{F}_s(r) P_X(ds) \leq \int_B \tilde{F}_s(r') P_X(ds)$$

for every $B \in \mathcal{S}$. This implies that

$$\tilde{F}_s(r) \leq \tilde{F}_s(r') \qquad \text{a.e.} \quad (P_X) \quad \text{on} \quad S,$$

i.e.

$$P_X(S(r, r')) = 1.$$

Set

$$S_1 = \bigcap_{r < r'} S(r, r').$$

Then S_1 is also a Borel subset of S with P_X-measure 1, because the right hand side is a countable intersection.

Since $\tilde{F}_s(r)$ is increasing in r for every $s \in S_1$, the function

$$(3) \qquad F_s(\xi) = \lim_{r \downarrow \xi} \tilde{F}_s(r), \qquad \xi \in R^1$$

is well-defined and it is non-decreasing and right continuous in ξ for every $s \in S_1$. It is obvious that

$$0 \leq F_s(\xi) \leq 1 \quad \text{for every} \quad \xi \in R^1 \quad \text{and} \quad s \in S_1.$$

For $s \in S - S_1$ define $F_s(\xi)$ to be any fixed increasing right continuous function with $0 \leq F_s(\xi) \leq 1$, for example

$$F_s(\xi) = 0 \quad (\xi < 0) \quad \text{and} \quad = 1 \quad (\xi \geq 0).$$

It is easy to see that $F_s(\xi)$ is Borel measurable in $s \in S$ for every $\xi \in R^1$. Keeping $P_X(S_1) = 1$ in mind we can derive

$$(2'') \qquad \int_B F_s(\xi) P_X(ds) = P((-\infty, \xi] \cap X^{-1}(B)), \quad B \in \mathcal{S}$$

from (2') and (3).

Since $F_s(\xi)$ is increasing and right continuous in ξ and $0 \leq F_s(\xi) \leq 1$, we can find a unique measure ν_s on $\mathcal{B}^1$ with $\nu_s(R^1) = F_s(\infty) - F_s(-\infty) \leq 1$ by applying 2.2 Theorem 5 to the function

$$G_s(\xi) = (F_s(\xi) - F_s(-\infty))/(F_s(\infty) - F_s(-\infty))$$

in case the denominator is positive and by setting $\nu_s = 0$ otherwise.

We will verify the following for every $A \in \mathcal{B}^1$.

(4) $\nu_s(A)$ is Borel measurable in s and

$$\int_B \nu_s(A) P_X(ds) = P(A \cap X^{-1}(B)), \quad B \in \mathcal{S}.$$

Let $\mathcal{D}$ be the class of all A's that satisfy this condition. $\mathcal{D}$ is obviously a Dynkin class. Since $\nu_s(A) = F_s(\xi)$ for $A = (-\infty, \xi]$, (4) holds for such an A. This implies that $\mathcal{D}$ includes the class

$$\mathcal{O} = \{(-\infty, \xi], \ \xi \in R^1\}.$$

Since $\mathcal{O}$ generates the σ-algebra $\mathcal{B}^1$, we have $\mathcal{D} \supset \mathcal{B}^1$ by Dynkin's lemma. This proves (4) for every $A \in \mathcal{B}^1$.

Since $\Omega \in \mathcal{B}^1$ and so $\mathcal{F} = \mathcal{B}^1 \cap \Omega \subset \mathcal{B}^1$, (4) holds for every $A \in \mathcal{F}$, in particular for $A = \Omega$. Therefore

$$\int_S \nu_s(\Omega) P_X(ds) = P(\Omega \cap X^{-1}(S)) = P(\Omega) = 1 = P_X(S).$$

Since $0 \leq \nu_s(\Omega) \leq \nu_s(R^1) \leq 1$, the set

$$S_2 = \{s \in S : \nu_s(\Omega) = 1\}$$

is a Borel subset of S with P_X-measure 1. It is obvious that the restriction $\nu_s|\Omega$ is a probability measure on $\mathcal{F}$ for $s \in S_2$.

~~Define μ_s by~~
Set

$$P^{X=s} = \begin{cases} \text{the Lebesgue extension of } \nu_s|\Omega \text{ if } s \in S_2, \\ \text{any fixed regular probability measure on } (\Omega, \mathcal{F}) \\ \qquad \text{if } s \in S - S_2. \end{cases}$$

Since $P_X(S_2) = 1$, we can easily verify (CP.1), (CP.2) and (CP.3) for $P^{X=s}$, $s \in S$.

Proof of (b). Since $(\Omega, \mathfrak{F})$ is a standard Borel space, we can find a countable multiplicative class $\mathfrak{A}$ of sets in $\mathfrak{F}$ which generates the σ-algebra $\mathfrak{F}$. Since we have

$$\int_B P_1^{X=s}(A)P_X(ds) = P(A \cap X^{-1}(B)) = \int_B P_2^{X=s}(A)P_X(ds)$$

for every $B \in \mathcal{S}$,

$$P_1^{X=s}(A) = P_2^{X=s}(A) \quad \text{a.e.} \quad (P_X) \quad \text{on} \quad S.$$

Denote the exceptional s-set by N_A and set $N = \cup_{A\in\mathfrak{A}} N_A$. Then $P_X(N) = 0$. For every $s \in S - N$ we have

$$P_1^{X=s}(A) = P_2^{X=s}(A) \quad \text{for every} \quad A \in \mathfrak{A}.$$

By 2.2 Theorem 2 we have $P_1^{X=s} = P_2^{X=s}$ for $s \in S - N$. This completes the proof of (b).

The family $\{P^{X=s}\}_s$ is uniquely determined not in the naive sense but in the sense "up to P_X-measure 0." In other words there are many versions of $\{P^{X=s}\}_s$ any two of which coincide for almost every (P_X) $s \in S$ where the exceptional s-set depends on the two versions. Therefore for a particular value of s, say s_0, $P^{X=s_0}$ has no unique meaning in general. But we have the following theorem.

Theorem 1. Suppose that $P(X=s_0) > 0$. Then $P^{X=s_0}$ is uniquely determined independently of the version. More precisely we have

(a) $A \in \mathfrak{M}(P^{X=s_0})$ if and only if $A \cap (X=s_0) \in \mathfrak{M}(P)$

and

(b) $P^{X=s_0}(A) = P(A \cap X^{-1}(s_0))/P(X^{-1}(s_0))$, $A \in \mathfrak{M}(P^{X=s_0})$.

Proof: Write μ for $P^{X=s_0}$ and define $\nu(A)$ to be the right hand side of (b) if and only if $A \cap (X=s_0) \in \mathfrak{M}(P)$. It is obvious that both μ and ν are regular probability measures on $(\Omega, \mathfrak{F})$. The assertion of our theorem is equivalent to

$$\mu = \nu$$

for every version of $\{P^{X=s}\}$. Setting $B = \{s_0\}$ in the equation in (CP.3), we have

$$P^{X=s_0}(A) P_X(\{s_0\}) = P(A \cap X^{-1}(s_0))$$

i.e.

$$\mu(A) = \nu(A)$$

for $A \in \mathfrak{F}$. Since μ and ν are both regular, we have $\mu=\nu$.

For $P^E(P(E) > 0)$ we have mentioned that $\mathfrak{M}(P^E)$ includes $\mathfrak{M}(P)$. But we cannot expect $\mathfrak{M}(P^{X=s}) \supset \mathfrak{M}(P)$ even if we allow an exceptional P_X-null s-set. The maximum we can say is the following.

<u>Theorem 2</u>. For every $A \in \mathfrak{M}(P)$ we have

$$A \in \mathfrak{M}(P^{X=s}) \qquad \text{a.e. } (P_X) \text{ on } S$$

where the exceptional s-set depends on A and on the version of $\{P^{X=s}\}_s$. For every version we have

$$\mathfrak{M}(P^{X=s}) \supset \mathfrak{M}(P)$$

for every s such that $P(X=s) > 0$.

<u>Proof</u>: Since $A \in \mathfrak{M}(P)$, we have $A_1, A_2 \in \mathfrak{F}$ such that

$$A_1 \subset A \subset A_2 \quad \text{and} \quad P(A_2 - A_1) = 0$$

by regularity of P. Then we have

$$\int_S P^{X=s}(A_2-A_1)P_X(ds) = P((A_2-A_1) \cap X^{-1}(S)) = 0.$$

by (CP.3) and so

$$P^{X=s}(A_2-A_1) = 0 \qquad \text{a.e. } (P_X) \text{ on } S.$$

This completes the proof of the first part. The second part is obvious by Theorem 1.

<u>Theorem 3</u>. For every $A \in \mathfrak{M}(P)$, $P^{X=s}(A)$ is P_X-measurable in s and we have

$$\int_B P^{X=s}(A)P_X(ds) = P(A \cap X^{-1}(B))$$

for every $B \in \mathfrak{M}(P_X)$. Setting $B = S$, we have

$$P(A) = \int_S P^{X=s}(A)P_X(ds),$$

which shows that P is determined by P_X and $\{P^{X=s}\}_s$.

Remark. $P^{X=s}(A)$ is defined a.e. (P_X) on S for $A \in \mathfrak{M}(P)$ by Theorem 2 and therefore the statement of this theorem is meaningful.

Proof of the theorem. Take A_1 and A_2 in the proof of Theorem 2. Then

$$P^{X=s}(A) = P^{X=s}(A_1) = P^{X=s}(A_2) \text{ a.e. } (P_X) \text{ on } S.$$

Since $P^{X=s}(A_1)$ is Borel measurable in s and so P_X-measurable in s, $P^{X=s}(A)$ is P_X-measurable in s and

$$\int_B P^{X=s}(A)P_X(ds) = \int_B P^{X=s}(A_1)P_X(ds) = P(A_1 \cap X^{-1}(B))$$

$$= P(A \cap X^{-1}(B)) - P((A-A_1) \cap X^{-1}(B)) = P(A \cap X^{-1}(B)).$$

Therefore our theorem holds for $B \in \mathcal{B}$. If $B \in \mathfrak{M}(P_X)$, we have $B_1 \in \mathcal{B}$ such that $B_1 \subset B$ and $P_X(B-B_1) = 0$. Then we have

$$\int_B P^{X=s}(A)P_X(ds) = \int_{B_1} P^{X=s}(A)P_X(ds) = P(A \cap X^{-1}(B_1))$$

$$= P(A \cap X^{-1}(B)) - P(A \cap X^{-1}(B-B_1)) = P(A \cap X^{-1}(B)),$$

because

$$0 \leq P(A \cap X^{-1}(B-B_1)) \leq P(X^{-1}(B-B_1)) = P_X(B-B_1) = 0.$$

For $P^E (P(E)>0)$ we have also mentioned that P^E is concentrated on E. This holds for $P^{X=s}$ for almost every (P_X) $s \in S$, namely

<u>Theorem 4</u>.

$$P^{X=s}(X^{-1}(s)) = 1 \qquad \text{a.e. } (P_X) \text{ on } S,$$

where the exceptional s-set depends on the version of $\{P^{X=s}\}_s$. Whatever version we may take, this equation holds for every $s \in S$ with $P(X=s) > 0$.

<u>Proof</u>. Since $(S, \mathcal{S})$ is a standard Borel space, it is Borel isomorphic with a Borel subset of $(R^1, \mathcal{B}^1)$. Therefore we can find a sequence of classes of Borel subsets of S:

$$\mathcal{E}_n = \{E_{n1}, E_{n2}, \ldots, E_{n,k(n)}\}, \quad n = 1, 2, \ldots$$

satisfying the following conditions.

(D.1) $\mathcal{E}_n$ is a division of S, i.e. $E_{n1}, E_{n2}, \ldots, E_{n,k(n)}$ are disjoint and $S = \cup_i E_{ni}$.

(D.2) For every n; $\mathcal{E}_{n+1}$ is a <u>subdivision</u> of $\mathcal{E}_n$, i.e. every member of $\mathcal{E}_n$ is the union of a number of members in $\mathcal{E}_{n+1}$.

(D.3) $\{\mathcal{E}_n\}_n$ is <u>separating</u>, i.e. for every pair of two different points s and t in S we have $\mathcal{E}_n$ such that s and t belong to different members of $\mathcal{E}_n$.

It is easy to see that if $E_n(s)$ denotes the member of $\mathcal{E}_n$ that contains s, then we have

(5) $E_1(s) \supset E_2(s) \supset \cdots \to \{s\}.$

Let E be any Borel subset of S. Then

$$\int_E P^{X=s}(X^{-1}(E))P_X(ds) = P(X^{-1}(E) \cap X^{-1}(E)) = P_X(E).$$

Since (the value of) the integrand is in $[0,1]$ for every s, we have

(6) $P^{X=s}(X^{-1}(E)) = 1$ for almost every (P_X) $s \in E$.

Let $N(E)$ denote the exceptional s-set. Then $P_X(N(E)) = 0$. Write N_{ni} for $N(E_{ni})$ and N for $\bigcup_{n,i} N_{ni}$. Then $P_X(N) = 0$.

Setting $E = E_{ni}$ in (6) we have

$$P^{X=s}(X^{-1}(E_{ni})) = 1, \quad s \in E_{ni} - N_{ni}$$

for $n = 1,2,\ldots$ and $i = 1,2,\ldots k(n)$. Suppose that $s \in S-N$. Then $s \in E_n(s)-N$ where $E_n(s)$ is E_{ni} for some i depending on s. Therefore

$$s \in E_{ni} - N \subset E_{ni} - N_{ni}.$$

Thus we have

$$P^{X=s}[X^{-1}(E_n(s))] = 1.$$

Since $X^{-1}(E_n(s)) \downarrow X^{-1}(s)$ $(n \to \infty)$ by (5), we have

$$P^{X=s}(X^{-1}(s)) = 1 \quad \text{for} \quad s \in S-N.$$

This completes the proof of the first part of our theorem. The second part is obvious by Theorem 1.

Suppose that $P(E) > 0$. Let $X(\omega)$ be the indicator of E. $X(\omega)$ is obviously a real random variable and

$$E = \{\omega : X(\omega) = 1\}.$$

Since $P(X=1) = P(E) > 0$, Theorem 1 shows that

$$P^{X=1} = P^E.$$

Therefore it is enough to discuss only the properties of $\{P^{X=s}\}_s$ as we will do below.

Take an arbitrary version of $\{P^{X=s}\}_s$ and fix it. Then we have a family of standard probability spaces

$$(\Omega, \mathfrak{F}, P^{X=s}), \quad s \in S.$$

On each probability space we can define a random variable, the probability law $(P^{X=s})_Y$ of a random variable Y, the expectation $E^{X=s}(Z)$ of a real or complex random variable Z and so on. $(P^{X=s})_Y$ is called the <u>conditional proability law</u> of Y <u>under</u> $X=s$ and $E^{X=s}(Z)$ (if exists) the <u>conditional expectation</u> of Z <u>under</u> $X=s$. Since $\mathfrak{M}(P^{X=s})$ varies with s, a $(T,\mathfrak{T})$-valued random variable on $(\Omega,\mathfrak{F},P^{X=s})$ for a value of s is not always so for another value of s. But we have the following.

<u>Theorem 5</u>. If $Y(\omega)$ is a $(T,\mathfrak{T})$-value random variable on $(\Omega,\mathfrak{F},P)$, then it is so on $(\Omega,\mathfrak{F},P^{X=s})$ for almost every (P_X) $s \in S$. In particular, if $Y(\omega)$ is Borel measurable, i.e. measurable $\mathfrak{F}/\mathfrak{B}^{\mathfrak{T}}$, then $Y(\omega)$ is a $(T,\mathfrak{T})$-valued random variable for every $s \in S$.

Proof. The second part is obvious because $\mathcal{M}(P^{X=s}) \supset \mathcal{F}$ for every s. To prove the first part, consider a map $Y_1 : \Omega \to T$ measurable $\mathcal{F}/\mathcal{J}$ such that

(7) $$Y_1(\omega) = Y(\omega) \qquad \text{a.e. } (P) \text{ on } \Omega;$$

the existence of such a Y_1 was shown in the first part of the proof of 2.3 Theorem 3. Let N be the exceptional ω-set. Then

$$\int_S P^{X=s}(N) P_X(ds) = P(N \cap X^{-1}(s)) = P(N) = 0$$

by Theorem 3. Therefore

$$P^{X=s}(N) = 0 \quad \text{a.e. } (P_X) \quad \text{on } S.$$

Let M be the exceptional s-set. Then

$$Y_1(\omega) = Y(\omega) \qquad \text{a.e. } (P^{X=s}) \quad \text{on } \Omega$$

for $s \in S-M$. As $Y_1(\omega)$ is a $(T, \mathcal{J})$-valued random variable on $(\Omega, \mathcal{F}, P^{X=s})$ by the second part of our theorem, so is $Y(\omega)$ for $s \in S-M$. This completes the proof of the first part.

Theorem 6. Let $Z(\omega)$ be a real or complex random variable on $(\Omega, \mathcal{F}, P)$. If $Z \geq 0$, then

(8) $$\int_B E^{X=s}(Z) P_X(ds) = E(Z, X^{-1}(B)) \text{ for } B \in \mathcal{M}(P_X).$$

In general, $Z \in L^1(\Omega, \mathcal{F}, P)$ if and only if

(i) $Z \in L^1(\Omega, \mathcal{F}, P^{X=s})$ for almost every (P_X) $s \in S$

and

(ii) $E^{X=s}(Z)$ (as a function of s) $\in L^1(S, \mathcal{S}, P_X)$.

(8) holds for $Z \in L^1(\Omega, \mathcal{F}, P)$.

Proof. If Z is the indicator of a set $A \in \mathfrak{M}(P)$, then (8) holds by Theorem 3. By taking linear combinations and monotone increasing limit, we can easily verify (8). If Z takes real or complex numbers, then we have

$$\int_S E^{X=s}(|Z|) P_X(ds) = E(|Z|)$$

This implies the second assertion. Writing $Z = Z^+ - Z^-$ $(Z^+ = Z \vee 0, Z^- = (-Z)^+)$ we can verify (8) for Z real. Writing $Z = \mathcal{R}Z + i\mathcal{J}Z$, we can verify (8) for Z complex.

Let $X(\omega)$ and $Y(\omega)$ be random variables with values in $(S, \mathcal{S})$ and $(T, \mathcal{T})$ respectively. Then $Z(\omega) = (X(\omega), Y(\omega))$ is a random variable with values in $(S \times T, \mathcal{S} \otimes \mathcal{T})$. We want to prove

$$(P^{X=s})_Y(dt)P_X(ds) = P_{(X,Y)}(d(s,t)).$$

The precise meaning of this symbolic formula is as follows.

<u>Theorem 7</u>. Let $f(s,t)$ be a $P_{(X,Y)}$ - integrable real or complex function on $S \times T$. Then

(a) $f(s,t)$ is $(P^{X=s})_Y$-integrable in $t \in T$ for almost every (P_X) $s \in S$.

(b) $\int_T f(s,t)(P^{X=s})_Y(dt)$ is P_X-integrable, and

(c) $\int_S \int_T f(s,t)(P^{X=s})_Y(dt)P_X(ds) = \int_{S \times T} f(s,t)P_{(X,Y)}(d(s,t))$.

<u>Proof</u>. The proof is similar to the proof of Fubini's theorem on product measures. If $f(s,t)$ is the indicator $1_{B \times C}$ of $B \times C$, $B \in \mathcal{S}$, $C \in \mathcal{T}$, we can prove (a), (b) and (c) by setting $A = Y^{-1}(C)$ in Theorem 3. The class $\mathcal{D}$ of all sets E for which (a), (b) and (c) holds for $f = 1_E$ is a Dynkin class including the class $\mathfrak{A}$ of all sets $E = B \times C$, $B \in \mathcal{S}$, $C \in \mathcal{T}$. Since $\mathfrak{A}$ generates the σ-algebra $\mathcal{S} \otimes \mathcal{T}$, we have $\mathcal{D} \supset \mathcal{S} \otimes \mathcal{T}$ by Dynkin's lemma. The rest of the proof can be carried out in the same way as the proof of Fubini's theorem.

2.5. Iteration of conditioning

Let $X(\omega)$ be an $(S,\mathcal{S})$-valued random variable on $(\Omega,\mathfrak{F},P)$ and $\{P^{X=s}\}_s$ the conditional probability measures relative to X defined in the previous section. Let $Y(\omega)$ be a $(T,\mathfrak{T})$-valued random variable in $(\Omega,\mathfrak{F},P)$. Then $Y(\omega)$ is also a $(T,\mathfrak{T})$-random variable on $(\Omega,\mathfrak{F},P^{X=s})$ for every $s \in S-N$, N being a P_X-null subset of S, by virtue of 2.4 Theorem 5. If $Y(\omega)$ is Borel measurable i.e. measurable $\mathfrak{F}/\mathfrak{T}$, then we have $N = \phi$. Even if $Y(\omega)$ is not Borel measurable but only P_X-measurable, we can take a version of $\{P^{X=s}\}_s$ for which $N = \phi$. To prove this, take a Borel subset M of S with $M \supset N$ and $P_X(M) = 0$ and then replace $P^{X=s}$ for $s \in M$ by the regular probability measure δ_{ω_0} on $(\Omega,\mathfrak{F})$ concentrated at any fixed point $\omega_0 \in \Omega$. From now on we will take this version of $\{P^{X=s}\}_s$. Then we can again define the conditional probability measures $(P^{X=s})^{Y=t}$, $t \in T$ for each $s \in S$, so that we obtain a doubly indexed family

$$(P^{X=s})^{Y=t}, \quad s \in S, \; t \in T.$$

This family has many versions according to the choice of a version of $\{P^{X=s}\}_s$ and the choice of a version of $\{(P^{X=s})^{Y=t}\}_t$ for each $s \in S$. Not every version of $\{(P^{X=s})^{Y=t}\}_{s,t}$ is useful. The only useful version is a jointly Borel measurable version, i.e. a version for which $(P^{X=s})^{Y=t}(A)$ is Borel measurable as a function of $(s,t) \in (S \times T, \mathcal{S} \otimes \mathfrak{T})$ for every $A \in \mathfrak{F}$.

Theorem 1.

(a) There exists a jointly Borel measurable version of $\{(P^{X=s})^{Y=t}\}_{s,t}$.

(b) Let $\{(P_1^{X=s})^{Y=t}\}_{s,t}$ and $\{(P_2^{X=0})^{Y=t}\}_{s,t}$ be jointly Borel measurable versions of $\{(P^{X=s})^{Y=t}\}_{s,t}$. Then

$$(P_1^{X=s})^{Y=t} = (P_2^{X=s})^{Y=t}$$

for almost every $(P_{(X,Y)})$ $(s,t) \in S \times T$.

Proof of (a). Take any version of $\{P^{(X,Y)=(s,t)}\}_{(s,t)}$ and any version of $\{P^{X=s}\}_s$. ~~such that Y(ω) is $P^{X=s}$-measurable for every s.~~ Suppose that $A \in \mathfrak{F}$, $B \in \mathcal{S}$ and $C \in \mathcal{T}$. Applying 2.4 Theorem 7 to

$$f(s,t) = 1_B(s)1_C(t)P^{(X,Y)=(s,t)}(A),$$

we have

$$\int_B \int_C P^{(X,Y)=(s,t)}(A)(P^{X=s})_Y(dt)P_X(ds)$$

$$= \int_{B\times C} P^{(X,Y)=(s,t)}(A)P_{(X,Y)}(d(s,t))$$

$$= P(A \cap (X,Y)^{-1}(B \times C))$$

$$= P(A \cap Y^{-1}(C) \cap X^{-1}(B))$$

$$= \int_B P^{X=s}(A \cap Y^{-1}(C))P_X(ds).$$

For A and C fixed, this is true for every $B \in \mathcal{S}$ and so for every $B \in \mathfrak{M}(P_X)$. By 2.4 Theorem 7(b) and Theorem 3, we have

$$(1) \quad \int_C P^{(X,Y)=(s,t)}(A)(P^{X=s})_Y(dt) = P^{X=s}(A \cap Y^{-1}(C))$$

for almost every (P_X) $s \in S$. Let N be the exceptional s-set. Since $P_X(N) = 0$, we have a Borel subset M of S such that $N \subset M$ and $P_X(M) = 0$. Then (1) holds for $s \in S-M$. Since M depends on A and C, write it as $M(A,C)$. Since $(\Omega,\mathcal{F})$ is standard, we have a countable multiplicative family $\{A_n\} \subset \mathcal{F}$ which generates the σ-algebra $\mathcal{F}$. Set $M(C) = \bigcup_n M(A_n, C)$. Then $P_X(M(C)) = 0$ and (1) holds for $A = A_n, n=1,2,\ldots$ if $s \in S-M(C)$. Using Dynkin's lemma as we did several times, we can prove (1) for every $A \in \mathcal{F}$ if $s \in S-M(C)$. Noticing that $(T,\mathcal{T})$ is standard and using Dynkin's lemma again, we have a P_X-null s-set M outside of which (1) holds for every $A \in \mathcal{F}$ and every $C \in \mathcal{T}$. Now set

$$\mu_s = \begin{cases} P^{X=s} & \text{for } s \in S-M. \\ \text{the regular probability measure } \delta_{\omega_0} \text{ on } (\Omega,\mathcal{F}) & \\ \text{concentrated at } \omega_0 & \text{for } s \in M. \end{cases}$$

and

$$\mu_{s,t} = \begin{cases} P^{(X,Y)=(s,t)} & \text{for } s \in S-M,\ t \in T \\ \delta_{\omega_0} & \text{for } s \in M,\ t \in T. \end{cases}$$

Then $\mu_s(A)$ is Borel measurable in s for every $A \in \mathcal{F}$ and

$$\int_B \mu_s(A) P_X(ds) = \int_B P^{X=s}(A) P_X(ds) = P(A \cap X^{-1}(B))$$

by $P_X(M) = 0$. It is obvious that $Y(\omega)$ is μ_s-measurable for every s. $\mu_{s,t}(A)$ is Borel measurable in (s,t) because $P^{(X,Y)=(s,t)}$ is so. We can verify

$$\int_B \mu_{s,t}(A)(\mu_s)_Y(dt) = \mu_s(A \cap Y^{-1}(C)), \quad A \in \mathcal{F},\ C \in \mathcal{G}$$

by (1) for $s \in S-M$ and by trivial computation for $s \in M$. Thus $\{\mu_{s,t}\}_{s,t}$ is a jointly measurable version of $\{(P^{X=s})^{Y=t}\}_{s,t}$.

<u>Proof of (b)</u>. For any jointly measurable version of $\{(P^{X=s})^{Y=t}\}_{s,t}$ we have

$$\int_{B \times C} (P^{X=s})^{Y=t}(A)\ P_{(X,Y)}(d(s,t))$$

$$= \int_B \int_C (P^{X=s})^{Y=t}(A)(P^{X=s})_Y(dt)P_X(ds)$$

$$= \int_B P^{X=s}(A \cap Y^{-1}(C))P_X(ds)$$

$$= P(A \cap Y^{-1}(C) \cap X^{-1}(B))$$

$$= P(A \cap (X,Y)^{-1}(B \times C))$$

for $A \in \mathcal{F}$ by applying <u>2.4</u> Theorem 7 to

$$f(s,t) = 1_B(s)1_C(t)(P^{X=s})^{Y=t}(A).$$

Therefore we have

$$\int_E (P_1^{X=s})^{Y=t}(A)\ P_{(X,Y)}(d(s,t))$$

$$= \int_E (P_2^{X=s})^{Y=t}(A)\ P_{(X,Y)}(d(s,t))$$

for $E = B \times C$, $B \in \mathcal{S}$, $C \in \mathcal{T}$. The class of all such sets E is a multiplicative class generating the σ-algebra $\mathcal{S} \otimes \mathcal{T}$. Therefore the above equality holds for every $E \in \mathcal{S} \otimes \mathcal{T}$ by Dynkin's lemma. This implies the conclusion of (b). *Now use a Dynkin lemma argument again to show that the null set can be chosen independently of s.*

From now on we consider only jointly Borel measurable versions of $\{(P^{X=s})^{Y=t}\}_{s,t}$ unless stated otherwise.

<u>Theorem 2</u>. $\{(P^{X=s})^{Y=t}\}_{s,t}$ is a version of $\{P^{(X,Y)=(s,t)}\}_{(s,t)}$.

<u>Proof</u>. Since $(P^{X=s})^{Y=t}(A)$ is Borel measurable in (s,t) for every $A \in \mathcal{F}$, we have

$$\int_{B \times C} (P^{X=s})^{Y=t}(A)\ P_{(X,Y)}(d(s,t))$$

$$= P(A \cap (X,Y)^{-1}(B \times C)),\quad B \in \mathcal{S},\ C \in \mathcal{T},$$

as we have shown in the proof of Theorem 1(b). Using Dynkin's lemma again, we have

$$\int_{E} (P^{X=s})^{Y=t}(A)\ P_{(X,Y)}(d(s,t)) = P(A \cap (X,Y)^{-1}(E))$$

for every $E \in \mathcal{S} \otimes \mathcal{T}$. This completes the proof.

Theorem 2 means that

$$(P^{X=s})^{Y=t} = P^{(X,Y)=(s,t)} \text{ a.e. } (P_{(X,Y)}) \text{ on } S \times T.$$

If $X(\omega)$, $Y(\omega)$, and $Z(\omega)$ are random variables with values in $(S, \mathcal{S})$, $(T, \mathcal{T})$ and $(U, \mathcal{U})$ respectively, then the following conditional probability measures are equal to each other a.e. $(P_{(X,Y,Z)})$ on $S \times T \times U$:

$$P^{(X,Y,Z)=(s,t,u)}, \quad ((P^{X=s})^{Y=t})^{Z=u},$$

$$(P^{(X,Y)=(s,t)})^{Z=u}, \quad (P^{X=s})^{(Y,Z)=(t,u)}.$$

2.6. The conditional probability measure relative to a σ-algebra.

Let $(\Omega, \mathcal{F}, P)$ be a standard probability space. A σ-algebra on Ω included by $\mathfrak{M}(P)$ is called a sub-σ-algebra of $\mathfrak{M}(P)$. The system $\underline{\underline{B}}$ of all sub-σ-algebras of $\mathfrak{M}(P)$ is a semi-ordered system with respect to the set theoretical inclusion relation. Let $\underline{\underline{C}}$ be a subsystem of $\underline{\underline{B}}$. The greatest lower bound of $\underline{\underline{C}}$, $\bigwedge_{\mathcal{B}\in\underline{\underline{C}}} \mathcal{B}$ in notation, is the set theoretical intersection of all $\mathcal{B} \in \underline{\underline{C}}$ and the least upper bound of $\underline{\underline{C}}$, $\bigvee_{B\in\underline{\underline{C}}} \mathcal{B}$ in notation, is the σ-algebra generated by the set theoretical union of all $\mathcal{B} \in \underline{\underline{C}}$.

If $P(A \ominus B) = 0$, we say that A is <u>P-equivalent</u> to B, $A = B$ a.s. in notation. For a given $\mathcal{B} \in \underline{\underline{B}}$ the class of all A's that are equivalent to some $B \in \mathcal{B}$ is also a sub-σ-algebra of $\mathfrak{M}(P)$ and is denoted by $\overline{\mathcal{B}}$. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be sub-σ-algebras of $\mathfrak{M}(P)$. If for every $B_1 \in \mathcal{B}_1$ we have $B_2 \in \mathcal{B}_2$ with $B_1 = B_2$ a.s., then we write $\mathcal{B}_1 \subset \mathcal{B}_2$ a.s. If $\mathcal{B}_1 \subset \mathcal{B}_2$ a.s. and $\mathcal{B}_2 \subset \mathcal{B}_1$ a.s., then we write $\mathcal{B}_1 = \mathcal{B}_2$ a.s. and $\mathcal{B}_1$ is said to be <u>P-equivalent</u> to $\mathcal{B}_2$. $\mathcal{B}_1 \subset \mathcal{B}_2$ a.s. and $\mathcal{B}_1 = \mathcal{B}_2$ a.s. are equivalent to $\overline{\mathcal{B}}_1 \subset \overline{\mathcal{B}}_2$ and $\overline{\mathcal{B}}_1 = \overline{\mathcal{B}}_2$ respectively.

Let $\mathcal{B}$ be a sub-σ-algebra of $\mathfrak{M}(P)$. We can introduce an equivalence relation in Ω relative to $\mathcal{B}$ by writing $\omega_1 \sim \omega_2$ if and only if either $\omega_1, \omega_2 \in B$ or $\omega_1, \omega_2 \in B^c$ for every $B \in \mathcal{B}$. Each equivalence class is called an <u>atom</u> in $\mathcal{B}$. Every set $B \in \mathcal{B}$ is the union of a (finite or infinite) number of atoms in $\mathcal{B}$. An atom in $\mathcal{B}$ does not belong to $\mathcal{B}$ in general. Therefore the notion of atoms is not very useful for general sub-σ-algebras of $\mathfrak{M}(P)$.

Suppose that $\mathscr{B}$ is <u>countably generated</u>, i.e. generated by a countable subfamily of $\mathscr{B}$. In this case the following three conditions are equivalent.

(A.1) A is an atom in $\mathscr{B}$,

(A.2) A is a set $\in \mathscr{B}$ such that if $B \subset A$ and $B \in \mathscr{B}$ then $B = \phi$ or A.

(A.3) $A = \bigcap_n A'_n$, $A'_n = A_n$ or A_n^c, $n = 1,2,\ldots$ for any given countable subfamily $\{A_n\}$ that generates $\mathscr{B}$. In view of (A.2) and (A.3) every atom in $\mathscr{B}$ belong to $\mathscr{B}$ if $\mathscr{B}$ is countably generated.

Let us mention a remark on the property of being countably generated. Since $(\Omega,\mathscr{F})$ is standard, $\mathscr{F}$ is countably generated. But not every sub-σ-algebra of $\mathscr{F}$ is countably generated in general as we will show below. Therefore it is obvious that not every sub-σ-algebra of $\mathscr{M}(P)$ is countably generated in general. Suppose that $\Omega = [0,1]$ and $\mathscr{F} = \mathscr{B}^1 \cap [0,1]$. Then $(\Omega,\mathscr{F})$ is standard. Let λ denote the Lebesgue measure. Consider the class $\mathscr{B}$ that consists of all $B \in \mathscr{F}$ with $\lambda(B) = 0$ or 1. $\mathscr{B}$ is obviously a sub-σ-algebra of $\mathscr{F}$ but it is not countably generated. Suppose that it is generated by $\{A_n\} \subset \mathscr{B}$. Let A be the intersection of all A_n with $\lambda(A_n) = 1$ minus the union of all A_n with $\lambda(A_n) = 0$. It is obvious that $A \in \mathscr{B}$ and $\lambda(A) = 1$. Let $\mathscr{B}_1$ be the class of all sets $\in \mathscr{B}$ that include A or are included by A^c. $\mathscr{B}_1$ is obviously a sub-σ-algebra of $\mathscr{B}$. Since $\mathscr{B}_1$ includes all A_n, $\mathscr{B}_1$ includes $\mathscr{B}$. Therefore $\mathscr{B}_1 = \mathscr{B}$. Since $\lambda(A) = 1$, we have a proper subset B of A such that $B \in \mathscr{F}$ and $\lambda(B) = 1$. Then we have $B \in \mathscr{B}$ and $B \notin \mathscr{B}_1$ in contradiction with $\mathscr{B}_1 = \mathscr{B}$.

If we want to prove only the fact that not every sub-σ-algebra of $\mathfrak{M}(\lambda)$ is countably generated, we can do it more easily by ~~the~~ a cardinal number argument. In fact $\mathfrak{M}(\lambda)$ itself is not countably generated, because the cardinal number of $\mathfrak{M}(\lambda)$ is at least (in fact equal to) $2^{\underline{c}}$ as $\mathfrak{M}(\lambda)$ includes all subsets of the Cantor set, while the cardinal number of every countably generated σ-algebra is at most $\underline{\underline{c}}$.

Theorem 1. For every sub-σ-algebra $\mathcal{B}$ of $\mathfrak{M}(P)$ we have a countably generated sub-σ-algebra $\mathcal{B}_1$ of $\mathfrak{M}(P)$ which is P-equivalent to $\mathcal{B}$.

Proof. As we have mentioned in 2.3 Theorem 4, $\mathfrak{M}(P)$ is separable with respect to $\rho(A,B) = P(A \ominus B)$ and so is $\mathcal{B}$. Let $\mathfrak{A}$ be a countable ρ-dense subset of $\mathcal{B}$ and $\mathcal{B}_1$ the σ-algebra generated by $\mathfrak{A}$. We want to prove that $\mathcal{B}_1 = \mathcal{B}$ a.s. It is obvious that $\mathcal{B}_1 \subset \mathcal{B}$. Therefore it is enough to prove that for every $B \in \mathcal{B}$ we can find $B_1 \in \mathcal{B}_1$ such that $P(B_1 \ominus B) = 0$. Take $A_n \in \mathfrak{A}$ such that $P(A_n \ominus B) < 2^{-n}$. Noticing the obvious facts

$$E \ominus \bigcup_n E_n \subset \bigcup_n (E \ominus E_n)$$

$$E \ominus \bigcap_n E_n \subset \bigcup_n (E \ominus E_n),$$

we can easily prove that $P(B \ominus B_1) = 0$ for $B_1 = \limsup_n A_n \in \mathcal{B}_1$.

Now we will define the conditional probability measure $P^{\mathcal{B},\omega}$ on $(\Omega, \mathcal{F})$ relative to a sub-σ-algebra $\mathcal{B}$ of $\mathfrak{M}(P)$. The intuitive meaning of $P^{\mathcal{B},\omega}$ is given by

$$P^{\mathcal{B},\omega}(A) = \lim_{\substack{B \downarrow \alpha(\omega) \\ B \in \mathcal{B}}} P^B(A) = \lim_{\substack{B \downarrow \alpha(\omega) \\ B \in \mathcal{B}}} \frac{P(A \cap B)}{P(B)}$$

where $\alpha(\omega)$ is the atom that includes ω, but the rigorous definition of $P^{\mathcal{B},\omega}$ is given by the following conditions.

(CP.1') $P^{\mathcal{B},\omega}$ is a regular probability measure on $(\Omega, \mathcal{F})$ for each $\omega \in \Omega$,

(CP.2') The map $\omega \to P^{\mathcal{B},\omega}(A)$ is measurable $\mathcal{B}/\mathcal{B}^1$ for each $A \in \mathcal{F}$,

(CP.3') $\int_B P^{\mathcal{B},\omega}(A)P(d\omega) = P(A \cap B)$, $A \in \mathcal{F}$, $B \in \mathcal{B}$. We will write $P^{\mathcal{B}}$ for $P^{\mathcal{B},\omega}$ if there is no possibility of confusion.

$P^{\mathcal{B},\omega}$ is well-defined, because we can prove the following statements in the same way as we proved (a) and (b) in Section 2.4.

(a') There exists such a $P^{\mathcal{B},\omega}$.

(b') If we have two such $P^{\mathcal{B},\omega}$, say $P_1^{\mathcal{B},\omega}$ and $P_2^{\mathcal{B},\omega}$, then

$$P_1^{\mathcal{B},\omega} = P_2^{\mathcal{B},\omega} \quad \text{a.s.}$$

There are many versions of $P^{\mathcal{B},\omega}$ which are equivalent to each other in the sense of (b').

<u>Theorem 2</u>. $P^{\overline{\mathcal{B}},\omega} = P^{\mathcal{B},\omega}$ a.s.

<u>Proof</u>. The map $\omega \to P^{\mathcal{B},\omega}(A)$ $(A \in \mathcal{F})$ is measurable $\overline{\mathcal{B}}/\mathcal{B}^1$ by (CP.2') and $\overline{\mathcal{B}} \supset \mathcal{B}$. It follows from (CP.3') that (CP.3') remains to hold even if we replace $\mathcal{B}$ by $\overline{\mathcal{B}}$. Therefore $P^{\mathcal{B},\omega}$ is a version of $P^{\overline{\mathcal{B}},\omega}$. Thus our theorem follows from (b').

Since "$\mathcal{B}_1 = \mathcal{B}_2$ a.s." is equivalent to "$\overline{\mathcal{B}}_1 = \overline{\mathcal{B}}_2$", we have the following by Theorem 2.

Theorem 3. If $\mathcal{B}_1$ and $\mathcal{B}_2$ are P-equivalent, then

$$P^{\mathcal{B}_1,\omega} = P^{\mathcal{B}_2,\omega} \quad \text{a.s.}$$

As an immediate result from Theorems 1 and 3, we have the following.

Theorem 4. For every sub-σ-algebra $\mathcal{B}$ of $\mathcal{M}(P)$ we have a countably generated sub-σ-algebra $\mathcal{B}_1$ of $\mathcal{M}(P)$ such that $P^{\mathcal{B},\omega} = P^{\mathcal{B}_1,\omega}$ a.s.

Similarly to 2.4 Theorem 2 and 3 we can prove the following.

Theorem 5. Suppose that $A \in \mathcal{M}(P)$. Then

(i) $A \in \mathcal{M}(P^{\mathcal{B},\omega})$ a.s.,

(ii) $P^{\mathcal{B},\omega}(A)$ is P-measurable in ω,

(iii) $\int_B P^{\mathcal{B},\omega}(A)P(d\omega) = P(A \cap B),\ B \in \mathcal{B}$.

By virtue of Theorem 4 we can restrict ourselves to the countably generated σ-algebras in discussing the conditional probability measures. Keeping this in mind we have a theorem for $P^{\mathcal{B},\omega}$ corresponding to 2.4 Theorem 4.

Theorem 6. If $\mathcal{B}$ is countably generated, then

$$P^{\mathcal{B},\omega}(\alpha(\omega)) = 1 \quad \text{a.s.}$$

where $\alpha(\omega)$ is the atom in $\mathcal{B}$ that contains ω.

Proof. The idea of the proof is the same as that of 2.4 Theorem 4. Since we have

$$\int_A P^{\mathcal{B},\omega}(A)P(dw) = P(A \cap A) = P(A)$$

for $A \in \overline{\mathcal{B}}$ ~~$\mathfrak{M}(P)$~~ by Theorem 5, we have

$$P^{\mathcal{B},\omega}(A) = 1 \quad \text{for} \quad \omega \in A-N(A)$$

where $N(A)$ is a subset of A with $P(N(A)) = 0$.

Let $\{A_n\}$ be a countable subfamily of $\mathcal{B}$ generating $\mathcal{B}$ and set

$$A(\epsilon_1,\epsilon_2,\ldots,\epsilon_n) = \bigcap_{i=1}^{n} A_i^{\epsilon_i} \text{ and } N(\epsilon_1,\epsilon_2,\ldots,\epsilon_n) = N(A(\epsilon_1,\epsilon_2,\ldots,\epsilon_n))$$

where $\epsilon_i = 0$ or 1, $A_i^0 = A_i^c$ and $A_i^1 = A_i$, $i = 1,2,\ldots$.
Let N be the union of all $N(\epsilon_1,\ldots\epsilon_n)$, $n = 1,2,\ldots$ $\epsilon_i = 0$ or 1 $(i = 1,2,\ldots)$. Then $P(N) = 0$.
For n fixed,

$$A(\epsilon_1,\epsilon_2,\ldots,\epsilon_n), \quad \epsilon_1,\epsilon_2,\ldots,\epsilon_n = 0 \text{ or } 1$$

are disjoint and the union of all such sets is Ω. Let $\alpha_n(\omega)$ be the set among those sets that contains ω. Then

$$\alpha_1(\omega) \supset \alpha_2(\omega) \supset \ldots \to \alpha(\omega).$$

It is easy to see that

$$P^{\mathcal{B},\omega}(\alpha_n(\omega)) = 1 \quad \text{for} \quad \omega \in \Omega-N.$$

Therefore $P^{\mathcal{B},\omega}(\alpha(\omega)) = \lim_n P^{\mathcal{B},\omega}(\alpha_n(\omega)) = 1$, $\omega \in \Omega - N$. This completes the proof.

Let $X(\omega)$ be an $(S,\mathcal{S})$-valued random variable and $\sigma[X]$ the σ-algebra generated by X i.e.

$$\sigma[X] = X^{-1}(\mathcal{S}) = \{X^{-1}(B);\ B \in \mathcal{S}\}.$$

Then we define the conditional probability measure $P^{\sigma[X],\omega}$. We have defined $\{P^{X=s}\}_s$ in Section 2.4. The following theorem that can be proved easily will connect these two notions with each other.

Theorem 7. $$P^{X=s}\Big|_{s=X(\omega)} = P^{\sigma[X],\omega} \quad \text{a.s.}$$

where the left hand side means $P^{X=s}$ evaluated at $s = X(\omega)$.

It is customary to write P^X for $P^{\sigma[X]}$. Because of the following theorem we can interpret $P^{\mathcal{B}}$ in terms of P^X.

Theorem 8. Let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{M}(P)$. Then there exists a real random variable $X(\omega)$ such that

$$\mathcal{B} = \sigma[X] \text{ a.s. and so } P^{\mathcal{B}} = P^X \text{ a.s.}$$

Proof. By Theorem 1 we have a countably generated sub-σ-algebra $\mathcal{B}_1$ of $\mathcal{M}(P)$ which is P-equivalent to $\mathcal{B}$. It is enough to prove that we have a real random variable X such that $\mathcal{B}_1 = \sigma[X]$. Let $\{A_n\}$ generate $\mathcal{B}_1$ and $e_n(\omega)$ the indicator of A_n, $n = 1,2,\ldots$.
Set

$$X(\omega) = \sum_{n=1}^{\infty} \frac{2\, e_n(\omega)}{3^n} .$$

It is obvious that X is a real random variable. As every e_n is measurable $\mathcal{B}_1/\mathcal{B}^1$, so is X. This implies $\sigma[X] \subset \mathcal{B}_1$. Since $X(\omega)$ belongs to the Cantor set, $X(\omega)$ has a unique triadic expansion and so $e_n(\omega)$ is determined by $X(\omega)$ as follows:

$$2e_n(\omega) = [3^n X(\omega)] - 3[3^{n-1}X(\omega)],\ n = 1,2,\ldots ,$$

$[\alpha]$ denoting the maximum integer $\leq \alpha$. Since $\alpha \to [\alpha]$ is Borel measurable from R^1 into itself, we have

$$A_n = \{\omega : e_n(\omega) = 1\} = \{\omega : X(\omega) \in E\}$$

with some $E \in \mathcal{B}^1$. Therefore $A_n \in \sigma[X]$ for every n. This implies $\mathcal{B}_1 \subset \sigma[X]$ and so $\mathcal{B}_1 = \sigma[X]$.

As in the case of $P^{X=s}$ we can define probabilistic concepts such as random variables, the probability law $(P^{\mathfrak{B},\omega})_Y$ of a random variable Y, the expectation $E^{\mathfrak{B},\omega}(Z)$ of a real or complex random variable Z etc. on the probability space $(\Omega, \mathfrak{F}, P^{\mathfrak{B},\omega})$. $(P^{\mathfrak{B},\omega})_Y$ and $E^{\mathfrak{B},\omega}(Z)$ are called the <u>conditional probability law</u> of Y under $\mathfrak{B}$ and the <u>conditional expectation</u> of Z under $\mathfrak{B}$ respectively.

Analogously to <u>2.4</u> Theorems 5 and 6 we have the following theorems.

<u>Theorem 9</u>. If Y is a $(T, \mathfrak{T})$-valued random variable on $(\Omega, \mathfrak{F}, P)$, then it is so on $(\Omega, \mathfrak{F}, P^{\mathfrak{B}})$ a.s. In particular, if $Y(\omega)$ is measurable $\mathfrak{F}/\mathfrak{T}$, then $Y(\omega)$ is a $(T, \mathfrak{T})$-valued random variable on $(\Omega, \mathfrak{F}, P^{\mathfrak{B}})$ ~~for every $\omega \in \Omega$.~~ everywhere on Ω.

<u>Theorem 10</u>. Let Z be a real or complex random variable on $(\Omega, \mathfrak{F}, P)$. If $Z \geq 0$, then

$$(1) \qquad E(E^{\mathfrak{B},\omega}(Z), B) = E(Z, B) \quad \text{for} \quad B \in \bar{\mathfrak{B}}.$$

In general, $Z \in L^1(\Omega, \mathfrak{F}, P)$ if and only if

(i) $Z \in L^1(\Omega, \mathfrak{F}, P^{\mathfrak{B},\omega})$ a.s.

and

(ii) $E^{\mathfrak{B},\omega}(Z)$ (as a function of ω) $\in L^1(\Omega, \mathfrak{F}, P)$.

The formula (1) holds for $Z \in L^1(\Omega, \mathfrak{F}, P)$.

Suppose that $\mathfrak{B}_1 \subset \mathfrak{B}_2$. We want to discuss the relation between $P^{\mathfrak{B}_1}$ and $P^{\mathfrak{B}_2}$.

Theorem 11.

(i) If $A \in \mathcal{F}$, $E^{\mathcal{B}_1}[P^{\mathcal{B}_2}(A)]$ is well-defined for every ω. This is a probability measure on $\mathcal{F}$ as a function of A. The Lebesgue extension of this measure gives a version of $P^{\mathcal{B}_1}$.

(ii) If $A \in \mathcal{M}(P)$, then

$$E^{\mathcal{B}_1}[P^{\mathcal{B}_2}(A)] = P^{\mathcal{B}_1}(A) \quad \text{a.s.}$$

(iii) If $Z \in L^1(\Omega, \mathcal{F}, P)$, then

$$E^{\mathcal{B}_1}[E^{\mathcal{B}_2}(Z)] = E^{\mathcal{B}_1}(Z).$$

Proof. (i) is obvious by the definition. If $A \in \mathcal{M}(P)$, then $P^{\mathcal{B}_2}(A)$ is P-measurable. Therefore it is $P^{\mathcal{B}_1}$-measurable a.s. and

$$E[E^{\mathcal{B}_1}(P^{\mathcal{B}_2}(A)), B] = E(P^{\mathcal{B}_2}(A), B), \quad B \in \overline{\mathcal{B}}_1$$

by Theorem 10 and for $B \in \overline{\mathcal{B}}_1$ we have

$$\begin{aligned} E(P^{\mathcal{B}_2}(A), B) &= P(A \cap B) \quad \text{by} \quad \overline{\mathcal{B}}_2 \supset \overline{\mathcal{B}}_1 \\ &= E(P^{\mathcal{B}_1}(A), B). \end{aligned}$$

Thus we have

$$E[E^{\mathcal{B}_1}(P^{\mathcal{B}_2}(A)), B] = E(P^{\mathcal{B}_1}(A), B) \quad \text{for} \quad B \in \overline{\mathcal{B}}_1.$$

Since $E^{\mathcal{B}_1}(P^{\mathcal{B}_2}(A))$ and $P^{\mathcal{B}_1}(A)$ are both measurable $\overline{\mathcal{B}}_1/\mathcal{B}^1$, the above equation implies

$$E^{\mathcal{B}_1}(P^{\mathcal{B}_2}(A)) = P^{\mathcal{B}_1}(A) \quad \text{a.s.}$$

This proves (ii). The last assertion (iii) is obvious for Z = the indicator of $A \in \mathfrak{M}(P)$ by (ii). Using this we can prove (iii) for Z general by a routine method.

<u>Theorem 12</u>. $U = E^{\mathfrak{B}}(Z)$ a.s. if and only if

(i) U is measurable $\mathfrak{B}$

and

(ii) $E(U,B) = E(Z,B)$ for $B \in \mathfrak{B}$.

<u>Proof</u>. If $U = E^{\mathfrak{B}}(Z)$, then U satisfies (i) and (ii) by Theorem 10. The converse is obvious.