2. THE PROBLEM OF DESCRIPTION

Consider the following data.

Landings by Fishing Craft at Boston, Gloucester and New Bedford, 1948-1950.

(millions of pounds)

	Boston			Gloucester			New Bedford		
Variety	1948	1949	1950	1948	1949	1950	1948	1949	1950
Cod	34.7	28.6	24.4	7.7	7.0	6.0	6.3	4.1	5.1
Haddock	105.3	90.1	107.4	11.2	8.5	10.0	11.4	9.9	11.5
Mackerel	3.5	0.7	1.1	19.9	6.9	5.7	1.7	0.3	0.3
Flounder	9.2	8.1	10.0	2.7	7.6	7.4	41.0	33.9	29.4

Source: U.S. Bureau of the Census, Statistical Abstract of the United States: 1956 (Washington, 1956), p. 720.

A listing of the "ingredients" of this table might run as follows:-

a "universe of discourse" consisting of certain regions (ports), at certain time periods (1948-1950), at which certain resources (types of fish) appear; and

a measure of the quantities involved for each possible combination of place, time, and resource.

It turns out that this data can, in fact, be represented by a measure in the technical mathematical sense of the term.

Furthermore, this is true not only for the present example, but for ~~the great bulk~~ of statistical data in general -- population, births, migration, marriages, production, transportation, income, wealth, ~~and so on.~~

Even this rather sweeping statement understates the possibilities of describing the world in terms of measures. ~~For it turns out that~~ these kinds of data which cannot be represented <u>directly</u> as measures -- such ~~things~~ as prices, population densities, per-capita incomes -- are <u>derived</u> from underlying measures in ways which themselves are well-known parts of the theory.

This chapter will ~~be devoted to~~ carrying out this descriptive program. It will expound the concepts, terminology and basic theorems of measure theory as used in this book.[1] And it will build a unified apparatus for describing the world in terms of these concepts. The unity arises from the fact that the "universes of discourse" over which the various measures are defined are always built up from the three basic sets of Space, Time, and Resources, just as in the fish example above (but not always in so simple a fashion).

Of course, the fact that this apparatus <u>can</u> be constructed does not mean it is <u>useful</u> to do so. The rest of this book may be considered an argument for the proposition that it <u>is</u> useful.

2.1. Measure Theory, I

This section is a purely formal exposition: No real-world interpretations are offered. We shall in general follow the practice of omitting proofs of theorems when these are available in standard treatises.[2]

The concept of "measure" rests on three other concepts: (i) "sigma-field"; (ii) "the extended real numbers"; and (iii) "countable additivity". We discuss these in turn.

Sigma-Fields

We begin by recalling some notions from elementary set theory. Write $x \in A$ to indicate that x <u>belongs to</u>, or is a <u>member</u>, or <u>element</u>, or <u>point</u>, of the set A. $A \subseteq B$ (or $B \supseteq A$) indicates that A is a <u>subset</u> of B (that is, every element of A is an element of B). $A = B$ signifies that $A \subseteq B$ and $B \subseteq A$ (so that sets are considered equal iff they have the same members; we use "iff" to abbreviate "if and only if" throughout this work). $\{x_1, x_2, \ldots\}$ is the set whose elements are $x_1, x_2, \ldots$. $\{x\}$ is the set which has the single member x. $\emptyset$, the Norwegian letter "O" stands for the <u>empty set</u>, the set which has no members.

Let P be a certain property, and let the symbol

$$\{x \mid x \text{ has the property } P\}$$

stand for the set of all objects having the property P. Given two sets, A and B, their <u>intersection</u>, written $A \cap B$, is defined as the set of elements they have in common – that is,

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

The union of A and B, written $A \cup B$, is the set of elements which are in at least one of the two sets – that is,

$$A \cup B = \{x \mid x \in A \text{ or } x \in B \text{ (or both)}\}.$$

The complement of B with respect to A, written $A \backslash B$ (A slash B) is the set of elements which are in A but not in B:

$$A \backslash B = \{x \mid x \in A \text{ and } x \notin B\}$$

(A line drawn through any relation signifies that the corresponding proposition is not true: thus $A \not\subset B$, $A \neq B$, as well as $x \notin B$).

Now consider a set, $\mathcal{G}$, whose elements are themselves sets. For euphony, $\mathcal{G}$ will be called a class, or a collection, of sets, rather than a set of sets. (We shall follow the convention of using small letters for points, capital letters for sets of points, and script letters for classes of sets).

written $\cap\mathcal{G}$, the intersection of $\mathcal{G}$, is defined as the set of points common to all the members of $\mathcal{G}$:

$$\cap\mathcal{G} = \{x \mid x \in G \text{ for all } G \in \mathcal{G}\}$$

It is clear that if $\mathcal{G}$ has as members just the two sets A and B, this reduces to $A \cap B$. Similarly, one defines the union of $\mathcal{G}$, written $\cup\mathcal{G}$, as $\{x \mid x \in G \text{ for at least one } G \in \mathcal{G}\}$.

One has the basic distinction between sets with a finite and with an infinite number of members (finite and infinite sets, for short.) Another distinction of great importance is that between countable and uncountable sets. A set is countably infinite iff its members can be ticked off in an infinite sequence: x_1, x_2, x_3, ... (In other words, the set can be placed in 1 - 1 correspondence with the positive integers). Examples of countably infinite sets are: the set of positive integers, the set of all integers, the set of rational numbers, the set of lattice points in the plane (that is, the set of all points (m,n) where m,n are both integers). On the other hand, the set of rial numbers is not countably infinite.

A set is countable iff it is either finite, or countably infinite. The following result is used repeatedly in this book, generally without explicit mention.

Theorem: Let G be a countable collection of sets, each of which is itself countable; then $\bigcup G$ is countable.

We are now ready to define "sigma-field" (sometimes called "sigma-algebra"). Suppose one is given A, a set, and Σ, a collection of subsets of A.

Definition: Σ is a sigma-field (with universe set A) iff

(i) $\emptyset \in \Sigma$ and $A \in \Sigma$; and

(ii) if $E \in \Sigma$ and $F \in \Sigma$, then $E \setminus F \in \Sigma$; and

(iii) if G is a countable collection of members of Σ, then $\bigcup G$ and $\bigcap G$ are both members of Σ.

(One expresses (ii) by saying that Σ is closed under differences, and (iii) by saying that Σ is closed under countable unions and intersections.)

Definition: If Σ is a sigma-field with universe set A, the pair (A,Σ) is called a measurable space; the members of Σ are called measurable sets.

The definition of sigma-field given above is redundant, in the sense that some of the conditions follow from the others. To verify that a given collection of sets is a sigma-field, it is then useful to have a stripped-down criterion. First, if the universe set A is given, then $A \setminus B$ is simply called the complement of B.

Theorem: Let Σ be a collection of subsets of A. Σ is a sigma-field (with universe set A) iff $\emptyset \in \Sigma$, and Σ is closed under complementation and countable unions (that is, $E \in \Sigma$ implies that $A \setminus E \in \Sigma$, and $\cup G \in \Sigma$ for any countable collection G of sets belonging to Σ).

Thus we need to verify only half of conditions (i) and (iii), and a weakened form of (ii).

We now give some examples of sigma-fields.

(i) The collection of the two sets A, $\emptyset$ by themselves constitute a rather trivial sigma-field.

(ii) The collection of all subsets of A (including A and $\emptyset$) is a sigma-field.

(iii) For this example we first give some definitions which are quite important in their own right. Let $\mathcal{G}$ be a collection of subsets of $\underline{A}$.

Definitions: $\mathcal{G}$ is a covering (of $\underline{A}$) iff $A \subseteq \bigcup \mathcal{G}$; $\mathcal{G}$ is a packing iff no two members of $\mathcal{G}$ have a point in common (that is, if $\underline{G}_1$ and $\underline{G}_2$ belong to $\mathcal{G}$, then $\underline{G}_1 \cap \underline{G}_2 = \emptyset$); $\mathcal{G}$ is a partition iff it is both a packing and a covering.

These definitions may also be expressed as follows: $\mathcal{G}$ is a covering iff every point of $\underline{A}$ belongs to at least one member of $\mathcal{G}$; $\mathcal{G}$ is a packing iff every point of $\underline{A}$ belongs to at most one member of $\mathcal{G}$; $\mathcal{G}$ is a partition iff every point of $\underline{A}$ belongs to exactly one member of $\mathcal{G}$.

Now let $\mathcal{G}$ be a given partition of $\underline{A}$, and let Σ consist of all sets of the form $\bigcup \underline{F}$, where $\underline{F}$ ranges over all possible subclasses of $\mathcal{G}$. (Since $\bigcup \emptyset = \emptyset$, the empty set belongs to Σ). One may verify that Σ is then a sigma-field.

An important special case arises when $\mathcal{G}$ is a finite partition. If $\mathcal{G}$ has non-empty member sets, then the sigma-field Σ has 2^N member sets. It may be shown that all finite sigma-fields are of this form.

(iv) In the three examples above, one could give a simple property characterizing the measurable sets (that is, the members of Σ). Usually, however, this is not possible. Instead, one typically characterizes sigma-fields as being generated from a class of sets given in advance. We now turn to this important concept.

Let S be a non-empty collection of sigma-fields, all relative to the same universe set A. One may verify that ∩S is then itself a sigma-field relative to A. (Notice that S is a "third-order" construct: It is a set whose members are sets, the members of these in turn being sets; ∩S is then a "second-order" construct, a class of sets — which in fact turns out to be a sigma-field).

Definition: Let G be a collection of subsets of A; the sigma-field generated by G is ∩S, where S is the collection of all sigma-fields (relative to A) containing G as a subclass:

$$S = \{\Sigma \mid \Sigma \text{ is a sigma-field and } G \subseteq \Sigma\}.$$

As an example, let G be a countable partition; then the generated sigma-field is precisely Σ as constructed above under (iii). (To prove this, one shows first that Σ is indeed a sigma-field, and second that every member of Σ must belong to every sigma-field containing G; this second statement follows at once from the fact that sigma-fields are closed under countable unions).

One may say for short that the sigma-field generated by a collection G is the "smallest" sigma-field containing G. The discussion above explicates the concept and shows that indeed there is such a smallest sigma-field.

With the aid of this concept we may now define what is historically the granddaddy of all sigma-fields: the Borel field on the real line. Here the universe set is the real line, and the Borel field is simply the sigma-field generated by the class of finite open intervals -- that is, all sets of the form $\{x \mid a < x < b\}$, where a and b range over the real numbers.

This sigma-field is quite important, and it is useful to note that it may be generated by a variety of different collections. Besides the finite open intervals, it is generated by the class of sets of the form $\{x \mid x < b\}$, b ranging over the real numbers; also by the class of sets $\{x \mid x < a\}$, a ranging over the reals; furthermore, it is generated by any of these classes when a or b ranges merely over the rational numbers instead of the real numbers.[3] Finally, closed sets could be used in any of these cases instead of open sets (just substitute the weak inequality "$\leq$" for the strict inequality "$<$").

There is no direct way of characterizing the real Borel field, and so one must be satisfied to define it in terms of "generation". As we mentioned above, this situation is the rule, not the exception.

The Extended Real Numbers

We now turn to the second concept needed in the definition of measure: the extended real number system. The ordinary real number system is augmented by two "points of infinity", and the relation of order and the operations of arithmetic are then extended to these ideal points.

Definition: The extended real number system consists of the real numbers together with two new points, written "$+\infty$" and "$-\infty$"; $a \geq b$, $a + b$, $a \cdot b$, etc., retain their usual meanings when a and b are both real numbers; when one or both of these is $\pm\infty$, the order relation is extended as follows:

$$+\infty > a;\ a > -\infty;\ +\infty > -\infty, \text{ for any real number } a.$$

The operations of arithmetic are extended as follows:

addition:

$a + (+\infty) = (+\infty) + a = +\infty$; $a + (-\infty) = (-\infty) + a = -\infty$, for any real number a.

$(+\infty) + (+\infty) = (+\infty)$; $(-\infty) + (-\infty) = -\infty$.

(The expressions $(+\infty) + (-\infty)$, and $(-\infty) + (+\infty)$ are not defined, and are to be considered meaningless.)

negation:

$-(+\infty) = -\infty$; $-(-\infty) = +\infty$.

subtraction.

The rule: $a - b = a + (-b)$ defining subtraction in terms of addition and nigation.

multiplication.

If a is an extended real number >0, then

$$a \cdot (+\infty) = (+\infty) \cdot a = +\infty;\ a \cdot (-\infty) = (-\infty) \cdot a = -\infty.$$

If a is an extended real number <0, then

$$a \cdot (+\infty) = (+\infty) \cdot a = -\infty;\ a \cdot (-\infty) = (-\infty) \cdot a = +\infty.$$

$0 \cdot a = a \cdot 0 = 0$ for any extended real number a.

(end of definition).

All of these extensions are "natural", except perhaps the rule that zero times any extended real number, *including* $\pm\infty$, yields zero (which is useful in integration theory). The fact that addition (hence subtraction as well) is not always defined reveals a basic disparity between real and extended real arithmetic. (Division by $\pm\infty$ has also not been defined, but this is of no importance, since no occasion arises in this book where one would want such an operation).

The number "$+\infty$" will usually be abbreviated simply to "∞"[4]. The Borel field of the *extended* real numbers (or the extended real line, in geometric terminology) may be defined as the sigma-field generated by the class of sets of the form $\{x|x > a\}$, where a ranges over the real numbers. (These sets are now subsets of the extended real numbers, so that the number $+\infty$ belongs to all of them).

It may be verified that this sigma-field consists precisely of all sets which have any of the following four forms:

$$E,\ E \cup \{\infty\},\ E \cup \{-\infty\},\ E \cup \{\infty\} \cup \{-\infty\},$$

where E ranges over the Borel field of the *real* numbers.

Countable Additivity

First we recall the definition of "function". A **function** f with **domain** A and **values in** B, written $f: A \to B$, is a set of ordered pairs (a,b), where $a \in A$ and $b \in B$, each point of A being the first component of exactly one such pair. For each $a \in A$, the point of B thus associated with it is called the value of f at a, and is written f(a). The set

$$\{b \mid b = f(a) \text{ for at least one } a \in A\}$$

is called the range of f. It need not include all of B.

Given a measurable space, (A,Σ), consider a function $\mu: \Sigma \to$ non-negative extended real numbers. That is, μ assigns to each measurable set a value which is either a non-negative real number, or $+\infty$.

Definition: μ is finitely additive iff, for every pair of measurable sets, E, F, which are disjoint (that is, $E \cap F = \emptyset$),

$$\mu(E \cup F) = \mu(E) + \mu(F).$$

Theorem: Let $\mu: \Sigma \to$ non-negative extended real numbers be finitely additive. If $E \subsetneq F$, where E, F are measurable sets, then $\mu(E) \le \mu(F)$.

Proof: Since E, F are measurable, so is $F \setminus E$. Also E, $F \setminus E$ are disjoint, and $E \cup (F \setminus E) = F$. Hence, by (1), $\mu(E) + \mu(F \setminus E) = \mu(F)$. Since $\mu(F \setminus E) \ge 0$, it follows that $\mu(E) \le \mu(F)$. □

The property expressed by this theorem is called monotonicity, and implies that μ takes on its maximum value for the universe set A.

If μ is finitely additive, and E, F, G are three measurable sets no pair of which have a point in common (that is, the class consists of these three sets is a packing), then

$$\mu(E \cup F \cup G) = \mu(E) + \mu(F \cup G) = \mu(E) + \mu(F) + \mu(G).$$

It is clear by induction that a similar rule extends to any finite packing. However, we want to go further, and define such an additivity rule for all countable packings of measurable sets, not just for all finite packings.

Definition: Let μ take values in the non-negative extended real numbers, its domain being sigma-field Σ. μ is countably additive iff, for any countable packing of measurable sets, $\mathcal{G}$, we have

$$\mu(\cup\mathcal{G}) = \mu(G_1) + \mu(G_2) + \mu(G_3) + \ldots . \qquad (1)$$

(Here $G_1, G_2, G_3, \ldots$ is any enumeration of the members of $\mathcal{G}$ in a sequence, and the right-hand side of (1) is to be understood as the ordinary sum of an infinite series.)

The possibility that μ may take on the value ∞ causes no problems. If $\mu(G_n) = \infty$ for some n, then the right side of (1), hence the left side, equals ∞; if the partial sums on the right increase beyond any finite bound, then both sides again must equal ∞.

Measures

We now put all these concepts together.

Definition: A measure, μ, is a function

(i) whose domain is a sigma-field, Σ,

(ii) which takes values in the non-negative extended real numbers,

(iii) which is countably additive,

(iv) and for which $\mu(\emptyset) = 0$.

(~~There is exactly~~ one function ~~which~~ satisfies ~~conditions~~ (i), (ii) and (iii) but not ~~condition~~ (iv), ~~namely,~~ the function assigning the value ∞ to all sets. Thus (iv) serves only to exclude this rather trivial case; ~~cf.~~ the first example below).

Definition: The triple (A, Σ, μ), where A is the universe set, Σ a ~~sigma~~-field (relative to A), and μ a measure with domain Σ, is called a measure space (whereas the first two alone, without μ, constitute a measurable space~~, as we have mentioned~~).

~~The~~ The following result establishes a "continuity" property of sorts for measures.

Theorem: Let (A, Σ, μ) be a measure space, and let $G_1, G_2, \ldots,$ be a sequence of measurable sets ~~which~~ is increasing; ~~that is~~, $G_1 \subseteq G_2 \subseteq G_3 \subseteq \ldots$ then limit $\mu(G_n)$ as $n \to \infty$ is $\mu(\cup \mathcal{G})$, ($\mathcal{G}$ being the collection of all the G's).

If instead the sequence is decreasing ($G_1 \supseteq G_2 \supseteq \ldots$), and $\mu(G_K) < \infty$ for some K, then $\lim \mu(G_n) = \mu(\cap \mathcal{G})$.

(To prove the first part, partition $\cup \mathcal{G}$ into sets $G_n \setminus G_{n-1}$ and apply countable additivity; to prove the second part, take complements with respect to G_1 and apply the first part).

We finish this section by giving some examples of measures.

(i) Let $(\underline{A},\Sigma)$ be any measurable space; the function assigning the value zero to every member of Σ is a measure (the identically zero measure, which we will write simply as "$\emptyset$"); so is the function assigning the value ∞ to every non-empty member of Σ (the identically infinite measure, written "∞").

(ii) Let the sigma-field Σ be finite. As has been mentioned, Σ is then generated by a partition $\underline{G}$, and we may assume $\emptyset \notin \underline{G}$. Assign non-negative numbers arbitrarily to the members of $\underline{G}$; any member of Σ has a unique representation $\cup\underline{F}$, where $\underline{F} \subseteq \underline{G}$; assign to this set the value equal to the sum of the numbers assigned to members of $\underline{F}$. The result is a measure.

(iii) Let $(\underline{A},\Sigma)$ again be any measurable space. Define μ by:

$\mu(\underline{E})$ = number of points in $\underline{E}$, if $\underline{E}$ is finite, $\underline{E} \in \Sigma$;

$\mu(\underline{E}) = \infty$, if $\underline{E}$ is infinite, $\underline{E} \in \Sigma$.

μ is a measure, the enumeration measure.

(iv) Again let $(\underline{A},\Sigma)$ be arbitrary. Choose a fixed point $\underline{a}_o \in \underline{A}$, and define μ by:

$\mu(\underline{E}) = 1$ if $\underline{a}_o \in \underline{E}$ and $\underline{E} \in \Sigma$;

$\mu(\underline{E}) = 0$ if $\underline{a}_o \notin \underline{E}$ and $\underline{E} \in \Sigma$.

μ is a special kind of measure, an atomic measure (we shall discuss this more fully later). Note that we need not assume that the singleton set $\{\underline{a}_o\}$ is itself measurable.

(v) Our last example is the most famous of all measures. Let $\underline{A}$ be the real line, and let Σ be the Borel field on it. It may be shown that there is exactly one measure μ having the property that

$$\mu\{x|a < x < b\} = b - a$$

for all pairs of real numbers such that $a < b$; that is, μ assigns to any finite open interval its length. μ is known as (one-dimensional) Lebesgue measure.

2.2. Representation of the Real World by Measures: Preliminaries

Having at least defined the concept of measure, we now go on to real-world interpretations. (We shall follow the practice throughout this chapter of alternating formal development of measure theory with interpretations).

First, some general philosophical comments. We can start with some portion of the real world, and represent it in the language of some formal system. Or, we can start with some formal system and interpret (or "apply") the statements in it to refer to some part of the world. The first process clearly involve a severe abstraction (only a small fraction of fats about the world can be translated into the formal system).[5] The point is that the second process also involves an abstraction; it is not always possible to find a "fact" corresponding to every valid statement in the formal theory. One must then be satisfied with a partial interpretation of the formal system.

As an example, consider the representation of time by the real numbers. It is easy to interpret statements like: "$t_1 > t_2$", or "$t_1 - t_2 = t_3 - t_4$". But what facts correspond to the statements "t is irrational", or "every Cauchy sequence $\{t_n\}$ has a limit"? As far as facts are concerned, one could do just as well representing time by the rational numbers; however, the real numbers are more convenient.

In exactly the same way, while measure theory is a remarkably flexible and natural instrument for describing the world, one cannot expect every statement in it to correspond to a fact. The formal apparatus of the theory is designed for mathematical power and elegance, and as a result one might say the theory outruns what can be observed or measured in the real world.

What kinds of real-world facts are representable by measures? As a first example, think of the countries of the world as being identified with their territories. Consider the set of all locations on the surface of the Earth. (We idealize by thinking of each location as an extensionless point). The United States is a certain subset of these points, Switzerland is another subset, etc. Furthermore, no two of these subsets have a point in common.[6] Thus the collection of all countries is a packing. If we add to this collection the set consisting of the rest of the world (it will include the high seas, Antartica, etc.), we have a partition of the surface of the Earth. This partition generates a sigma-field, namely, all sets of the form $\cup F$, where F ranges over all subcollections of this partition.

Now choose any fixed date, and to the set $\cup F$ assign the number which is the total population of the territory $\cup F$ at this date. The function thus defined on the sigma-field is a measure. For, obviously, the total population of the union of two disjoint regions is the sum of the populations of the respective regions, so that the function is finitely additive. Furthermore, since we have a finite sigma-field, finite additivity is equivalent to countable additivity in this case. This proves we have a measure.

It is clear that any territorial magnitudes which have the finite additivity property can be represented as measures in exactly the same way as population. This includes territorial area, wealth, coal reserves, miles of highways, and (for a fixed time-interval) steel production, steel consumption, births, deaths, marriages, divorces, murders, PhD's granted, and innumerable others.[7] Note that measurement units can be quite varied - numbers of objects, mass, dollar value, acres, etc.

Statistical tables presenting data of this sort will not typically write out the entire measure. If the surface of the Earth is partitioned into, say, 130 nations plus rest of world, a complete table would have to assign a value to each of the 2^{131} members of the sigma-field. This is obviously impossible in practice, and also unnecessary, since if we are given the values for the 131 partition elements, the value for any other measurable set is given by the addition of the values of the appropriate subclass of partition elements. Thus in practice tables will just give values for the generating partition, plus perhaps a few other "marginal subtotals".

Any table of statistical data, if it can be represented by a measure at all, can be represented in the foregoing simple form, with the sigma-field generated by a finite partition.

For a second example, suppose one wanted to represent the concept "quantity of time". We represent time itself as usual by the real numbers, each number representing an "instant". We want to assign numbers to various sets of time-instants to measure the quantity of time embodied in that set. To an interval $\{t \mid a < t < b\}$ (where a, b are real numbers with $a < b$) is assigned the value $b - a$. Also, it seems reasonable that "quantity of time" should be at least finitely additive. This suggests that Lebesgue measure, on the Borel field of the real line, is an appropriate mathematical representation of this intuitive concept.

This simple example is quite instructive in illustrating how the requirements of intuition, and considerations of mathematical power, combine to suggest the appropriate representation. First, what is the appropriate class of subsets of the real line for which the assignment of a number representing "quantity of time" is to be considered meaningful? Intuition demands that it include all subsets for which, conceptually, an observation could be made (say by observing the angle through which a clock-hand turns) — hence certainly all finite intervals should be included. Mathematical elegance demands that it be a sigma-field. To satisfy both of these demands, it must include at least the entire Borel field.[8]

This requirement leads again to the "outrunning" of the facts by the theory. For what observation could confirm the statement: "The quantity of time embodied in the set of all rational time-instants is zero"?

Consider next the various additivity conditions. As suggested above, finite additivity has strong intuitive appeal. But measure theory comes into its own with the stronger requirement of countable additivity.[9] ~~Now~~ it may perhaps be contended that if a real-world magnitude is already finitely additive, it is intuitively plausible — or even demanded by intuition — that it be in fact countably additive. We know of no philosophical discussion of this issue.[9a] Since it seems to be of some importance, we offer some reflections ~~on it~~ in an appendix to this section.

Appendix on Additivity

Let μ be a function defined on a ~~sigma~~-field with range in the non-negative extended reals, representing some real-world data. We suppose ~~that~~ μ is finitely additive. There is then some plausibility for the view that it should be countably additive. But why stop at countable additivity? Doesn't intuition demand that a real-world-representing μ be uncountably additive?

Here we have run ahead of ourselves, since we have yet to give a definition of "uncountable additivity". First we need another concept.

Definition: Given a set of extended real numbers, E, the supremum of E is the smallest extended real number ~~which is~~ at least as large as every member of E.

For example, the supremum of the real numbers is ∞; the supremum of the negative numbers is 0. One of the advantages of the extended real ~~numbers~~ is that every subset of them has a supremum.

Now let I be an arbitrary non-empty set — finite or infinite, countable or uncountable. Let f be a function with domain I, and range in the non-negative extended real ~~numbers~~. Take any finite subset of I — say $\{i_1, i_2, \ldots, i_n\}$ — and from the sum

$$f(i_1) + f(i_2) + \ldots + f(i_n).$$

The summation of f is now defined as the supremum of the set of these sums formed by ranging over all possible finite subsets of I.

It is easy to verify that, if I is finite, the summation of f is ~~simply~~ its summation over I itself; if I is countable, summation of f is ~~simply~~ the limit of the series formed when the element of I are arranged in any sequence. The generalization, ~~then,~~ arises when I is uncountable.[10]

With the aid of these concepts we may now formulate ~~the~~

Definition: μ is uncountably additive iff, for any packing of measurable sets, $\mathcal{G}$, such that $\cup\mathcal{G}$ is measurable, the summation of μ on $\mathcal{G}$ equals $\mu(\cup\mathcal{G})$.

(Here $\mathcal{G}$ plays the role of the index set I above, and $\mu(G)$ with $G \in \mathcal{G}$ corresponds to $f(i)$ with $i \in I$; $\mathcal{G}$ need not be countable).

Uncountable implies countable additivity. Reverting to our argument above, it is hard to see why intuition should swallow countable additivity but strain at uncountable additivity.

~~The trouble is that~~ uncountable additivity appears to be too strong a condition for many purposes. For example, Lebesgue measure does not satisfy it. To see this, note that if the function f on the index set I is identically zero, ~~then~~ its summation equals zero. Since any set is the union of a packing of singleton sets, it follows that, if μ is uncountably additive and μ{x} = 0 for all singleton sets {x}, μ must be identically zero. But Lebesgue measure assigns value zero to each singleton set, and is not identically zero, hence it is not uncountably additive.

Possibly the difficulty can be resolved by going beyond the (standard) extended real-number system. In any case, we ~~shall in the remainder of this book naively~~ assume ~~that~~ our representations are countably, but not necessarily uncountably, additive ~~— that is,~~ we assume ~~that~~ they are measures in the ordinary sense of the term. This allows us to apply the great resources of measure theory to real-world problems. In this respect we ~~are~~ merely following in the footsteps of the great probabilists and statisticians.

2.3. Space, Time, and Resources

We can~~, and shall,~~ give many more examples of the representation of facts by measures. But we want to do more than this. We want a unified way of looking at the world so that all ~~of~~ these examples fall out naturally as specializations, and ~~so that~~ we do not need an ad hoc argument for each new case.

The claim is that this unified view can be built up from three basic sets: "Resources", the set of resource-types; "Space", the set of locations; and "Time", the set of instants. ~~As a menimonic,~~ we abbreviate these here, and throughout the book, as R, S, T, respectively.

In this section we describe these three sets. In later sections we put them together.

Time~~, as we have mentioned,~~ is well-represented by the real numbers (smaller numbers being prior to larger numbers in the temporal sense). The measurable subsets of T will always be taken to be the Borel field on the real line.

Definition: A subset of Time is called a period iff it is measurable.

Space is thought of most naturally as a ~~three~~-dimensional continuum. Sometimes~~, indeed,~~ it is more useful to identify Space with the surface of the Earth, since almost all human activity takes place in a thin film at the surface (even in the "space age"). In this connection one often makes two further idealizations. First, the Earth is taken to be a perfect sphere. Second, the spherical surface itself is flattened into a planar region, or even an infinite plane; thus three dimensions collapse to two.

When dealing with Space and Time jointly, one needs a convention about what locations at two different times are to be considered identical. We shall always assume that the Earth is at rest. This "geocentric" convention would not be made by a physicist, but for social science purposes it is by far the most convenient. Thus "Portugal" can be identified with the same subset of Space at different times, whereas it would be wandering about under any other convention.

Let us turn to the problem of defining an appropriate sigma-field on S. By "appropriate" we refer to the following somewhat vague desidetata. First, any subset of Space on which, conceptually, an observation could be made should be included. In particular, the various simple geometric figures -- (cubes and spheres in 3-space, squares and circles in the plane, etc.) -- should be included. On the other hand, one should not go much beyond the sigma-field generated by these, because there arise both mathematical and conceptual difficulties in defining measures on these very rich classes. On the real line the Borel field fits these specifications pretty well. There is a natural generalization to higher dimensional sets which serves much the same purpose.

Definition: On the plane the (2-dimensional) Borel field is the sigma-field generated by the class of open rectangles -- that is, by the sets of the form $\{(x,y) \mid a < x < b,\ c < y < d\}$, a, b, c, d, being real numbers.

Just as on the real line this sigma-field is generated by many other simple classes — for example, by the open discs, the sets of the form

$$\{(x,y) \mid (x-a)^2 + (y-b)^2 < c^2\}$$

a, b, c being real numbers; also by the closed rectangles or discs obtained by substituting "$\leq$" for "$<$". All ordinary geometric plane figures (thought of as including their boundaries) belong to this sigma-field.

Similarly, in 3-space the 3-dimensional Borel field is that generated by the "open prisms"

$$\{(x_1, x_2, x_3) \mid a_i < x_i < b_i,\ i = 1, 2, 3\},$$

and this sigma-field is also generated by the open discs

$$\{(x_1, x_2, x_3) \mid (x_1-a_1)^2 + (x_2-a_2)^2 + (x_3-a_3)^2 < b^2\}.$$

Finally, given a subset E of 3-space, such as the (idealized) surface of the Earth, the relative Borel field of E is the class of sets of the form

$$\{E \cap F \mid F \in \text{3-dimensional Borel field}\}.$$

For example, if we take a plane or a line embedded in 3-space, their relative Borel fields according to this definition may be shown to coincide exactly with the 2- and 1-dimensional Borel fields respectively.

These constructions give reasonably good solutions to the problem of the "appropriate" sigma-field on S. In some cases another choice may be better. In particular, it is often sufficient to work with a small subclass of the complete Borel field (as in our example of population distribution by country).[11]

In much of this book it turns out that the particular structure of Space (dimensionality, shape, etc.) is irrelevant. In this case it suffices to think of S as just some arbitrary measurable space. This gain in generality is important, because we often want to deal with Space of a highly "non-Euclidean" character, its structure determined by the irregularities of transportation cost and land quality.

For the remainder of this book, we use the term "region" in the following technical sense.

Definition: A subset of Space is called a region iff it is measurable.

We now turn to Resources, R. This has a much more complicated structure than Space or Time. Fortunately a great many results do not depend on a detailed knowledge of this structure. Also, there are certain conceptual problems tied up with R. We shall accordingly present a "naive" description of R in the rest of this section, reserving the discussion of difficulties for an appendix.

An object is identified by specifying where it is, when it is, and what kind of thing it is. Space is the set of possible answers to the first question, *Time* to the second, and we mean *Resources* to be the set of possible answers to the third.

The elements of the set of Resources are *types* of things rather than specific entities. Thus "water" is a resource (or rather, a set of resources, since it can be differentiated by temperature, pressure, purity, etc.) but any specific drop of water must be identified further by its position in Space and Time.

What types of things, then, are included in R? All possible types that are relevant for the problem in hand, and with as much fineness of distinction as is useful for the problem in hand. This will include natural resource types: soils, minerals, water, air, vegetation, animals. It will include manufactured commodities, crops, machinery, and structures. It will include sewage, garbage, trash, and junk. It will include all types of people, distinguished by sex, age, race, skills, beliefs, attitudes, tastes, personality, and any other relevant trait. It may even include such intangibles as light, sound, electricity and gravity.[12]

Two apparent difficulties may be cleared up at once. The first refers to resource-types which are non-existent. Should "unicorn" be included in R? Actually, it does no harm to include non-existents: as we shall see, existence is described by a *measure* placed on R, not by R itself. Second, can uniqueness or individuality be represented by a model which deals only in types? The answer is yes, provided the distinctions made in R are sufficiently fine. If one gives a very detailed description of a certain person-type, there will be at most one person at one time fitting that description - say George Washington at Noon, July 4, 1776.

One sometimes distinguishes between different resource-types and different varieties, or qualities, of the same resource type - say, minor variations of brand-name goods. From our present point of view, different varieties are also simply different types. We do not take account of the fact that a codfish is somehow more similar to a mackerel than it is to a cabbage.

Let us turn to the problem of finding an "appropriate" sigma-field for R. Following our previous approach, we should include all subsets on which, conceptually, a measurement could be taken. Thus "man", "fish", "water", "glove", "car" determine sets (the set of resource types which are men, fish, etc.) which should be measurable. (One could systematically go through the dictionary, and most nouns and adjectives would determine measurable subsets of R in the same way. The trouble is that most English words are more or less vague, and borderline cases arise: "Is this creature to be considered a fish or not?") Once the class of conceptually observable subsets of R is determined, the sigma-field generated by them would be the recommended sigma-field.

Unlike the case of Space and Time, where the Borel fields are the natural choices, the proper choice of sigma-field for Resources is still up in the air, as the paragraph above indicates. Fortunately, as we have mentioned, nothing in this

~~book~~ hinges on a detailed specification, and it is sufficient to suppose that R comes supplied with some σ-field, making it a measurable space.

Appendix on Resources

We discuss certain additional problems concerning the set of Resources. First is the problem of self-reference. Among the attributes of people will be their mental states — their beliefs, perceptions, thoughts~~, etc.~~ But to describe these, we must refer back to R ~~itself~~ (and to S and T and measures over these sets~~, etc.~~). Further complications arise if these mental states refer to still other mental states, and we can even get an infinite regress of the kind sometimes discussed in connection with strategy and games: "He thinks that I think that he thinks...". We take this point up again in the more general context of multilayered theories (see section 8.2~~, below~~).

Second is the problem of inclusiveness. When we take account of our limited information and the fact that there are "more things...than are dreamed of in our philosophy," anything smaller than the set of all possible resource types may be descriptively inadequate. But the concept of an all-inclusive R is not very clear and may even entail a logical contradiction.

Third is the problem of complex resources. It will generally not do, for example, to think of an entire river valley as a single resource-type. Instead, one thinks of it as a spatial configuration of resources: water, soil, trees, roads, houses, people, etc.[13] But the same reasoning applies to each of these smaller units. An automobile of a certain specific type is just a spatial configuration of steel, rubber, glass, paint, etc., and we could continue down to N_1 molecular-atomic level.

From the practical point of view, where one stops on this analysis depends on the size of the unit one is interested in. The physiologist may take a person to be a configuration of tissues, the biochemist may view him as a configuration of molecules. The solid scientist rarely has occasion to split people up spatially in this way.

Now, for N_1 uses to which we are putting the set R, to call something a resource-type which is actually heterogeneous spatial configuration involves distortion and the distortion is greater the larger the object is. The reason is that we use the concept of resource-location pair, (r,s), referring to a resource-type r located at a point s which is somewhat ill-defined if the "resource-type" is by its nature spread over a region of greater or lesser extent.

For the scale on which social science models typically operate, the distortion involved in treating people as resource types is negligible for the most part. The same is probably true for most ordinary commodities, although, as mentioned above, we would draw the line at resources as bg as entire river valleys.

This discussion raises the question, if we steadfastly refuse to admit spatial configurations as resource-types, what becomes of R? One is driven to resolve one configuration into its components, and these into further components, until presumably, one arrives at a small number of "elementary particles" out of which everything else is built up. Now from the practical point of view this procedure is absurd: One hardly expects atomic physics to be a prerequisite for social science. Nonetheless, there is the chance that a mathematically convenient theory can be built up by following this route, and we offer a few speculations as to what it would look like.

There is a general tendency for the number of kinds of things to become less as one descends the spatial hierarchy. (A great many different types of houses can be constructed by arranging one type of brick in different ways). Suppose, to make the theory as simple as possible, that everything ultimately reduces to just one kind of thing: "matter". The objects of everyday life would then be identified with certain distributions of matter over Space (or, possibly, Space and Time) - that is, with measures assigned to each region the quantity of matter in it, all of these regions forming a sigma-field relative to the universe set, which is the region which the object in question actually occupies.

There are two great virtues to such a theory. First, it avoids the distortions which arise on treating people and commodity-types as members of R. Second, it eliminates R itself (by reducing it to a single point) and thus simplifies things and avoids all the other difficulties connected with $\underline{R}$ which we have been discussing.

In the remainder of this book we shall for the most part ignore the issues raised in this appendix, and "naively" interpret $\underline{R}$ as outlined in the main body of this section. This is done partly for pedagogic reasons -- descriptively the theory runs closer to intuition -- but mainly because we have not yet arrived at satisfactory answers to the issues raised.

2.4. Measure Theory II

We return to pure mathematics in this section to define some concepts needed for further developments.

Restricted Measures

Let $(\underline{A},\Sigma,\mu)$ be a measure space. Just as with any other function we may consider the restriction of μ to a sub-domain of its domain Σ. That is, we take a subclass $\Sigma' \subseteq \Sigma$ and define a function μ' with domain Σ' by the rule

$$\mu'(\underline{E}) = \mu(\underline{E}), \text{ all } \underline{E} \in \Sigma'. \qquad (1)$$

The only special condition we insist on is that Σ' itself be a sigma-field (not necessarily relative to the original universe set $\underline{A}$). One states that Σ' is a sub-sigma-field of Σ.

It is then immediate that μ' is a measure; for the fact that it takes values in the non-negative extended real numbers and that it is countably additive, follows at once from its definition, (1).

Two special cases deserve mention.

Definition: μ' is an aggregation of μ iff $\underline{A} \in \Sigma'$

That is, A still remains the universe set, although Σ' is a "thinningout" of the original sigma-field. As an example, let A be the surface of a sphere and Σ the Borel field on A. Let G be a finite partition of A into Borel sets, and let Σ' be the sigma-field generated by G. (The distribution of population by countries fits this model; See section 2). It is clear why the term "aggregation" is used for this relation; While μ gives, say, the complete distribution of population, μ' just gives the distribution for entire countries.

Definition: Given measure space (A,Σ,μ), and $B \in \Sigma$, μ' is the restriction of μ to B iff μ' is the restriction of μ whose domain Σ' is the class of all measurable subsets of B.

One easily verifies that this Σ' is, indeed, a sigma-field, whose universe set, however, is B, not A.

As an example, again take the case of population distribution over the surface of the Earth. One may be interested only in the distribution within some particular region B, in which case one studies the restriction of μ to B. In general, the notion of restriction to B enables one to isolate particular objects, activities, or situations within the overall description.

Each different measurable subset B yields a different restriction.

Sometimes one is given, not the entire measure μ, but pieces or patches, each defined on the measurable subsets of some set B. If these patches cover the entire universe set A, the question arises: Can these patches be put together to yield a single measure on the entire measurable space? The following theorem gives the answer.

Theorem: ("patching theorem"). Given a measurable space (A,Σ), and, for each $n = 1, 2, \ldots$, a measure space (or "patch") (B_n, Σ_n, μ_n), satisfying the conditions:

(i) $B_n \in \Sigma$ for all n, and $\mathcal{B}$, the collection of all the B_n's, is a covering of A (i.e., $\cup\mathcal{B} = A$);

(ii) $\Sigma_n = \{E \mid E \subseteq B_n \text{ and } E \in \Sigma\}$, for all n;

(iii) the μ_n's are compatible, in the sense that, if $E \in (\Sigma_{n_1} \cap \Sigma_{n_2})$, then $\mu_{n_1}(E) = \mu_{n_2}(E)$, for all n_1, n_2, and E.

then there is exactly one measure μ on (A,Σ), such that μ_n is the restriction of μ to B_n, for all n.

Proof: First we prove there is at most one such μ. Let $B_1' = B_1$, $B_2' = B_2 \setminus B_1$, and, in general $B_n' = B_n \setminus (B_1 \cup \ldots \cup B_{n-1})$. Now suppose μ satisfies the conclusion of the theorem. For any $G \in \Sigma$,

$$\mu(G) = \sum_{n=1}^{\infty} \mu(G \cap B_n') = \sum_{n=1}^{\infty} \mu_n(G \cap B_n'). \qquad (2.4.2)$$

(The first equality in (2) arises from the facts that the sets $\{G \cap B_n' \mid n = 1, 2, \ldots\}$ are a packing whose union is G, and

μ is countably additive; the second equality arises from the facts that $G \cap B_n' \in \Sigma_n$, and ~~that~~ μ_n is the restriction of μ to Σ_n).

Since μ is explicitly determined by the μ_n's in ~~formula~~ (2), it is unique.

It remains to show that the μ defined by (2) does actually satisfy the theorem. First we show that for each n, μ_n is the restriction of μ to Σ_n. Let $G \in \Sigma_n$. For all $k > n$, $G \cap B_k' = \emptyset$. For all $k \leq n$, $\mu_k(G \cap B_k') = \mu_n(G \cap B_k')$, by conditions (ii) (iii). Hence, by (2),

$$\mu(G) = \sum_{k=1}^{n} \mu_n(G \cap B_k') = \mu_n(G \cap B_n) = \mu_n(G),$$

proving that μ_n is the restriction of μ to B_n.

Next, $\mu(\emptyset) = 0$, since $\mu_n(\emptyset) = 0$ for all n. It remains to show only that μ given by (2) is countably additive. Let $\mathcal{G} = \{G_m | m = 1, 2, \ldots\}$ be a measurable packing. Then

$$\mu(\cup \mathcal{G}) = \sum_{n=1}^{\infty} \mu_n((\cup \mathcal{G}) \cap B_n')$$

$$= \sum_{n=1}^{\infty} \left[\sum_{m=1}^{\infty} \mu_n(G_m \cap B_n')\right]$$

$$= \sum_{m=1}^{\infty} \left[\sum_{n=1}^{\infty} \mu_n(G_m \cap B_n')\right] = \sum_{m=1}^{\infty} \mu(G_m).$$

(Here the first and last equalities come from (2), the second from the countable additivity of each μ_n, and the reversal of summation order in the third equality from the fact that all summands are non-negative.) This proves that μ is countably additive.

The most important case of the patching theorem is where the class $\mathcal{B}$ is a partition of A. In this case we say that measure space (A, Σ, μ) is the direct sum of the measure spaces (B_n, Σ_n, μ_n), and write

$$A = B_1 \oplus B_2 \oplus \cdots ,$$

$$\Sigma = \Sigma_1 \oplus \Sigma_2 \oplus \cdots ,$$

$$\mu = \mu_1 \oplus \mu_2 \oplus \cdots ,$$

or, for short, $\mu = \oplus_n \mu_n$, etc.

In fact, it is easy to see that, given any countable collection of measure spaces, with disjoint universe sets, there is a unique way of combining them into a direct sum (see chapter 4, page for full definition). The direct sum should not be confused with the (ordinary) sum of measures, which is defined only for identical sigma-fields; see below).

Product Spaces

Consider any function with domain I and range in Y. If one is mainly interested in the range space, such a function may be written as

$$(y_i),\ i \in I,$$

and is referred to as a family of elements of Y, indexed by set I. (This is more general than the notion of "subset of elements of Y" since it allows for repetitions; the same Y-element may be assigned to more than one I-element). We distinguish families from subsets by using parentheses () instead of braces {}. The elements of Y themselves may be of any nature, numbers, functions, or sets, for example.

Consider the case where Y is a collection of sets - let us rewrite it as $\mathcal{G}$ to conform to our customary notation. We then have a family of sets indexed by I:

$$(G_i),\ i \in I,$$

where each $G_i \in \mathcal{G}$.

Definition: The cartesian product of the family (G_i), $i \in I$, is the set

$$\{(g_i),\ i \in I \mid g_i \in G_i \ \text{ for all } i \in I\}$$

That is, each member of the cartesian product is itself a family of elements of $\cup\mathcal{G}$ indexed by I; specifically, a family having the property that the element assigned to index i, g_i always belongs to the set G_i assigned to index i in the original family of sets (G_i), $i \in I$. The cartesian product will be denoted by $\Pi_{i \in I} G_i$, or even by ΠG_i if no confusion is possible. If all of the G_i's are identical (= G, say), the cartesian product is written G^I.

If the index set I is finite the cartesian product assumes a fairly simple form. For example, let I contain just two elements — say $I = \{1,2\}$ — and let the family be (A_1,B_2) (that is, set A is assigned to $i = 1$, and B to $i = 2$). The cartesian product of this family may be identified with the class of **ordered pairs**

$$\{(a,b) \mid a \in A, b \in B\}.$$

Here A and B may overlap, or even be identical. The cartesian product in this case is written $A \times B$.

Similarly, if the family of sets is (A_i), $i = 1, \ldots, n$, then the cartesian product may be identified with the set of **ordered n-tuples** $(a_1, \ldots, a_n)$, where $a_i \in A_i$ for all $i = 1, \ldots, n$, and these choices are made in all possible ways. This may be written $A_1 \times A_2 \times \ldots \times A_n$. Again, some or all of the A_i's may be identical. If all $A_i = A$, this may be written A^n.

As an example, let each of the A_i's be the real line. Then $A_1 \times \ldots \times A_n$ is simply n-space, the set of all n-tuples of real numbers.

Let us now introduce measure-theoretic concepts. Suppose we have a family of measurable spaces, $[(A_i,\Sigma_i)]$, $i \in I$. That is, for each index i there is given a sigma-field Σ_i with universe set A_i. There is a standard method for defining a sigma-field on the cartesian product ΠA_i. First we introduce a preliminary concept.

Definition: Subset E of the cartesian product ΠA_i is a rectangle iff there is a family of sets (E_i), $i \in I$ such that $E_i \subseteq A_i$ for all $i \in I$, and $E = \Pi E_i$.

As an example let A be the real line, and let E_1, E_2 be two finite intervals of real numbers. Then $E_1 \times E_2$ is literally a rectangle in the plane, the plane being, of course, the cartesian product $A \times A$. This is the origin of the abstract concept "rectangle".

Definition: Given the family $[(A_i, \Sigma_i)]$, $i \in I$, the product sigma-field is the sigma-field on the cartesian product ΠA_i generated by the class of all rectangles $E = \Pi E_i$ having the following properties:

(i) $E_i \in \Sigma_i$ for all i, and

(ii) $E_i = A_i$ for all i except for at most one index i_o.

(In condition (ii) the phrase "at most one" could be replaced by "at most a finite number of" or "at most a countable number of". That is, one can show that all three of these classes of rectangles generate the same sigma-field. It follows that if the index set I is countable, condition (ii) is trivial and may be dropped from the definition.) Rectangles satisfying condition (i) are called measurable.

The product sigma-field is denoted by $\Pi_{i \in I} \Sigma_i$, or $\Pi \Sigma_i$, and the resulting product measurable space is then $(\Pi A_i, \Pi \Sigma_i)$. If the (A_i, Σ_i) are identical for all I (= (A, Σ), say) this may be written $(\underline{A}^{\underline{I}}, \Sigma^{\underline{I}})$, or perhaps $(\underline{A}, \Sigma)^{\underline{I}}$. When $\underline{I}$ is finite, so that the family may be written $[(\underline{A}_{\underline{i}}, \Sigma_{\underline{i}})]$, $\underline{i} = 1, \ldots, \underline{n}$, the product sigma-field is written $\Sigma_1 \times \Sigma_2 \times \ldots \times \Sigma_{\underline{n}}$; and the resulting product space is then $(\underline{A}_1 \times \ldots \times \underline{A}_{\underline{n}}, \Sigma_1 \times \ldots \times \Sigma_{\underline{n}})$.[15]

Example. Let Σ be the Borel field on the real line. Then one may verify that $\Sigma \times \Sigma$ is simply the Borel field on the plane, and in fact this provides an alternative definiiion for that sigma-field. Similarly, $\Sigma \times \Sigma \times \Sigma$ is the Borel field in 3-space, and we may define the Borel field in $\underline{n}$-space (or even in arbitrary cartesian products of the real line with itself) in an analogous way.

Suppose one is given a measure space of the form $(\underline{A} \times \underline{B}, \Sigma' \times \Sigma'', \mu)$. That is, the product measurable space is built up from the two components $(\underline{A}, \Sigma')$ and $(\underline{B}, \Sigma'')$ in the manner just described, and a measure μ is given whose domain is the product sigma-field $\Sigma' \times \Sigma''$.

Definition: μ', the left marginal measure of μ, has domain Σ', and is given by

$$\mu'(\underline{E}) = \mu(\underline{E} \times \underline{B}),$$

all $\underline{E} \in \Sigma'$.

It is easily verified that μ' is indeed a measure. One may think of μ' as being constructed in two steps. First, one considers those members of $\Sigma' \times \Sigma''$ of the form $\underline{E} \times \underline{B}$, where $\underline{E} \in \Sigma'$. These form a sub-sigma-field, and μ restricted to this sub-domain is an aggregation, as defined above. Second, since the "right side" of all of the rectangles $\underline{E} \times \underline{B}$ is the same, we may regard μ as a function of its "left side" only; this yields μ'.

In the same way, the right marginal measure, μ'', with domain Σ'', is given by

$$\mu''(\underline{F}) = \mu(\underline{A} \times \underline{F}),$$

all $\underline{F} \in \Sigma''$.

As examples, take any cross-classification - say population classified by location and hair color, or shipments by origin and destination. If μ is the total distribution by numbers or mass, then the left marginal μ' will give the distribution of population by location alone, or of shipments by origin only. The right marginal will give population by hair color, or shipments by destination.

Statistical tables frequently give data for product spaces, and it is customary to give the marginal measures in addition to the original measure. (More accurately, to give the data for the generating partitions of the component measurable spaces. These are just the "marginal subtotals").

The "marginal" terminology appears in particular in probability theory.

Definition: A probability is a measure which assigns the value 1 to the universe set.

If μ in ($A \times B$, $\Sigma' \times \Sigma''$, μ) is a probability, one verifies immediately that μ' and μ'' are also probabilities — the left and right marginal probabilities, respectively.

Suppose one has an arbitrary product measurable space, ($\Pi_{i \in I} A_i$, $\Pi_{i \in I} \Sigma_i$). Let {I', I''} be a partition of the index set into two non-empty pieces. One may verify that the product space is the same as

$$\left[(\Pi_{i \in I'} A_i) \times (\Pi_{i \in I''} A_i),\ (\Pi_{i \in I'} \Sigma_i) \times (\Pi_{i \in I''} \Sigma_i)\right].$$

That is, we arrive at the same result by first taking the products over I' and I'', respectively, and then the product of these products.

Thus an arbitrary product space can be expressed as the product of two spaces in many ways. For any such factoring one can define left and right marginals exactly as above.

Measurable Functions

Let (A,Σ') and (B,Σ'') be two measurable spaces, and f a function with domain A and values in B.

(Note that, unlike measures, which assign values to subsets of A, f assigns values to individual points of A. It is customary to refer to the former type as set functions and the latter as point functions).

Definition: $\underline{f}$ is a <u>measurable function</u> (with respect to Σ', Σ'') iff, for all $E \in \Sigma''$, $\{a | f(a) \in E\} \in \Sigma'$.

The set $\{a | f(a) \in E\}$ is called the <u>inverse image</u> of E, so that the definition may be paraphrased: f is measurable iff the inverse image of every Σ''-measurable set is a Σ'-measurable set.

If there is no ambiguity, the reference to Σ', Σ'' may be omitted, and one simply writes: "f is measurable (or not)".

We give some examples:

(i) Let Σ' = <u>all</u> subsets of A. Then <u>any</u> function is measurable.

(ii) Let Σ'' consist of the two sets $\emptyset$, B. Then again any function is measurable (since the inverse image of B is A, and of $\emptyset$ is $\emptyset$).

(iii) Let f be a <u>constant</u> (that is, there is a $b_o \in B$ such that $f(a) = b_o$ for all $a \in A$). Then f is measurable. (Proof: If $b_o \in E$, the inverse image of E is A; if $b_o \notin E$, the inverse image is $\emptyset$).

(iv) Let $A = B$, and $\Sigma'' \subseteq \Sigma'$. Then the <u>identity</u> function (given by $f(a) = a$) is measurable. (Proof: The inverse image of any set is itself).

(v) Let (A,Σ') and (B,Σ'') both be the real line with Borel field. It may be shown that any <u>continuous</u> function is measurable.

(vi) Let $(A,\Sigma') = (B \times C, \Sigma'' \times \Sigma''')$. Let $f(b,c) = b$. Then f is measurable. (Proof: Let $E \in \Sigma''$; the inverse image of E is the set $E \times C$, which always belongs to $\Sigma'' \times \Sigma'''$.)

This last example has an important generalization. f is an example of a projection operator.

Definition: The projection from the cartesian product $\Pi_{i \in I} A_i$ to the i_o-th component space A_{i_o} is the function which assigns to the point (a_i), $i \in I$ the value a_{i_o}.

That is, it picks out the i_o-th "coordinate" of any element of the cartesian product. This function is written π_{i_o}. These projections are always measurable, the proof of this fact being a minor elaboration of that given under example (vi).

The following Theorem gives a very useful criterion for the measurability of a Function.

Theorem: Given measurable spaces (A, Σ), (B, Σ') and $f: A \to B$. Let $\mathcal{G}$ be a collection of sets which generates Σ'. Then f is measurable iff $\{a | f(a) \in G\} \in \Sigma$, for all $G \in \mathcal{G}$.

Proof: The "only if" statement is trivial. Conversely, let $\mathcal{F}$ be the class of all subsets of B whose inverse images are Σ-measurable. By assumption, $\mathcal{G} \subseteq \mathcal{F}$. If $E \in \mathcal{F}$, then $B \backslash E \in \mathcal{F}$; this follows from the fact that, since $\{a | f(a) \in E\}$ belongs to Σ, so does its complement $A \backslash \{a | f(a) \in E\} = \{a | f(a) \in B \backslash E\}$. Similarly, if $\mathcal{H} \subseteq \mathcal{F}$ and $\mathcal{H}$ is countable, then $\cup \mathcal{H} \in \mathcal{F}$. to see this, note that $\{a | f(a) \in H\} \in \Sigma$ for all $H \in \mathcal{H}$; hence the union of these sets over $H \in \mathcal{H}$ belongs to Σ; but this union is $\{a | f(a) \in \cup \mathcal{H}\}$, the inverse image of $\cup \mathcal{H}$. ~~Also $\emptyset \in \mathcal{F}$.~~

It follows that $\mathcal{F}$ is a sigma-field. Since it contains $\mathcal{G}$ which generates Σ', we must have $\mathcal{F} \supseteq \Sigma'$. Hence f is measurable. ▨

Let us apply this theorem to the case where B is the real or extended real numbers, and Σ' the corresponding Borel field. Σ' is generated by the class of sets $\{x|x > b\}$ where b ranges over the real numbers, and the same is true if ">" is replaced by any of the three signs "<", "≥", "≤". Hence to verify that some function f is measurable, it suffices to check that $\{a|f(a) > b\} \in \Sigma$ for all real b, or to do this with any of the other three signs in place of ">". (In fact, it suffices to check this for b rational).

This proof above gives a paradigm for proving general statements about all the members of a sigma-field: Prove the property for a generating class, and prove that the class possessing this property is closed under complements and countable unions. Another useful theorem proved in exactly this way is the following.

Theorem: ("measurable section theorem") Let $(A \times B, \Sigma' \times \Sigma")$ be a product space. For all $E \in (\Sigma' \times \Sigma")$, and for all $b \in B$, $\{a|(a,b) \in E\} \in \Sigma'$.

Proof: Let F be the class of subsets E of $A \times B$ having the property that, for all $b \in B$, $\{a|(a,b) \in E\} \in \Sigma'$. One verifies routinely that F is closed under complementation and countable unions. Next, consider the measurable rectangle $E' \times E"$. If $b \in E"$, then $\{a|(a,b) \in (E' \times E")\} = E'$; and if $b \notin E"$, then this set $= \emptyset$. Hence all such rectangles belong to F. But these generate $\Sigma' \times \Sigma"$. |||

Suppose we are given a function $f: A \times B \to C$. For a point $a_o \in A$ we define $f(a_o, \cdot)$ to be the function with domain B and range in C whose value at $b \in B$ is $f(a_o, b)$. This is the right a_o-section of f. Similarly, for $b_o \in B$, the left b_o-section of f, written $f(\cdot, b_o)$, is the function with domain A whose value at $a \in A$ is $f(a, b_o)$. Here the sets A and B may themselves be cartesian products.

Theorem: Given $(A \times B, \Sigma' \times \Sigma'')$ and (C,Σ), suppose ~~that~~ $f: A \times B \to C$ is measurable. Then all its left and right sections are measurable.

Proof: Consider any left section $f(\cdot, b_o)$. For any $E \in \Sigma$, the set $\{(a,b) | f(a,b) \in E\}$ belongs to $\Sigma' \times \Sigma''$, since f is measurable. Hence $\{a | f(a,b_o) \in E\} \in \Sigma'$, by the measurable section theorem above. But this set is the inverse image of E under $f(\cdot,b_o)$, hence the latter is measurable. The proof for right sections is similar.

Suppose one is given a measurable space (A,Σ) and a function $f: A \to B$. We ~~shall~~ use f and Σ to define a certain sigma-field on B.

Definition: The class, Σ', of all subsets $E \subseteq B$ having the property that $\{a | f(a) \in E\} \in \Sigma$ is called the sigma-field induced by f on B.

It is easily verified that Σ' is a sigma-field, and that f is measurable with respect to Σ,Σ'. In fact, Σ' may be characterized as the largest sigma-field on B such that f remains measurable with respect to Σ,Σ'.

This approach also works in reverse. Suppose this time that (B,Σ') is the measurable space, and again $f: A \to B$ is given.

Definition: The class, Σ, of all subsets of A of the form $\{a \mid f(a) \in E\}$, where E ranges over Σ', is called the sigma-field inversely induced by f on A.

Again, one verifies routinely that Σ is a sigma-field, and that f is measurable with respect to Σ,Σ'. In fact, Σ may be characterized as the smallest sigma-field on A such that f remains measurable with respect to Σ,Σ'.

Induction applies to measures as well as to sigma-fields.

Definition: Given measure space (A,Σ,μ), measurable space (B,Σ') and measurable function $f: A \to B$, the measure, μ', induced by f on Σ' is given by

$$\mu'(E) = \mu\{a \mid f(a) \in E\} \qquad (3)$$

all $E \in \Sigma'$.

One easily verifies that μ' is, in fact, a measure. As an example, take the product space $(A \times B, \Sigma \times \Sigma', \mu)$, the component space (B,Σ'), and let $f: A \times B \to B$ be the projection, given by $f(a,b) = b$. Then, for any $E \in \Sigma'$, we have, by (3),

$$\mu'(E) = \mu\{(a,b) \mid f(a,b) \in E\} = \mu(A \times E).$$

But this is precisely the definition of the right marginal measure so that this concept could have been defined as the measure induced on (B,Σ') by the projection of $A \times B$ on B. (Of course, the left marginal measure is that induced on (A,Σ) by the projection of $A \times B$ on A).

These inductions can be combined: Starting with a measure space (A,Σ,μ) and a function $f: A \to B$, one may first induce the sigma-field Σ' on B, and then the measure μ' on (B,Σ').

Measurability of functions is preserved under a grant variety of operations. We conclude by listing a few results of this type. The operations themselves are quite useful apart from any question of measurability.

Definition: Given sets A, B, C and functions $f: A \to B$ and $g: B \to C$, The composition of f and g, written $g \circ f$, is the function with domain A and range in C given by $(g \circ f)(a) = g(f(a))$.

Theorem: Given measurable spaces (A,Σ), (B,Σ'), (C,Σ''); if $f: A \to B$ and $g: B \to C$ are measurable, then so is $g \circ f$.

(To prove, merely note that the inverse image under $g \circ f$ is the inverse image under f of the inverse image under g).

We have already defined the supremum of a set of extended real numbers E, namely, as the smallest number not less than any $x \in E$. Now suppose we have a collection of functions F, all extended real-valued with common domain A.

Definition: The **supremum** of $\mathcal{F}$, written sup $\mathcal{F}$, is the function with domain A whose value at $a \in A$ is $\sup\{f(a) \mid f \in \mathcal{F}\}$.

Definition: Similarly, the **infinum** of a set of extended real numbers E, written inf E, is the largest number not greater than any $x \in E$. The **infinum** of collection $\mathcal{F}$ is the function whose value at $a \in A$ is $\inf\{f(a) \mid f \in \mathcal{F}\}$.

Theorem: Let (A,Σ) be a measurable space, and $\mathcal{F}$ a **countable** collection of extended real-valued functions with common domain A. If each $f \in \mathcal{F}$ is measurable, then sup $\mathcal{F}$ and inf $\mathcal{F}$ are measurable.

Given two functions $f,g: A \rightarrow$ reals, the **sum**, **difference** and **product** may be defined in the usual pointwise manner; e.g. $(f + g)(a) = f(a) + g(a)$. If f and g are measurable, so are their sum, difference and product.

Similarly, given a sequence of functions $f_n: A \rightarrow$ reals, $n = 1, 2, \ldots$, their **pointwise limit** (if it exists) is the function f whose value at $a \in A$ is $\lim_{n\rightarrow\infty} f_n(a)$. If all f_n are measurable, ~~then~~ f is measurable.

Given $f: A \rightarrow B$ and $E \subseteq A$, the **restriction** of f to E, written $f|E$, is the function with domain E ~~which~~ that coincides with f there. Given (A,Σ) and (B,Σ'), if f is measurable and $E \in \Sigma$, then $f|E$ is measurable.

Theorem: Given a measurable space (A,Σ), a family of measurable spaces $\{(B_i,\Sigma_i)\}_{i\in I}$ and a corresponding family of functions

$(f_i)_{i \in I}$, where $f_i: A \to B_i$, all i; let $g: A \to \Pi B_i$ be given by $g(a) = (f_i(a))_{i \in I}$. Then g is measurable iff all of the functions f_i, $i \in I$, are measurable.

2.5. Representation of the Real World by Measures: General Theory

We now have the tools to build a unified framework from the three basic sets, Resources, Space and Time.

Histories

Suppose we had a complete description of a person at some instant of his life. This will include his state - (height, weight, blood pressure, skills, attitudes, thoughts, etc. From our previous discussion of the set of Resources, this state may be identified with a point in R. It will also include his location, which is a point in S. We may identify this complete description, then, with a pair of points, one in R, one in S - and thus with a point in the cartesian product R × S.

This is for a single instant in Time. To give a complete lifetime picture of a person, we must repeat this procedure for such instant of his life. Suppose a person is born at time t_1 and dies at time t_2, so that he is alive in the interval $\{t | t_1 \leq t \leq t_2\}$.[16] A complete description would then represented by a function, whose domain is $\{t | t_1 \leq t \leq t_2\}$ and whose range is in R × S. Equivalently, it is represented by a pair of functions, both with domain $\{t | t_2 \leq t \leq t_2\}$, one with values in R and giving the person's state at each moment of his life, the other with values in S and giving his location at each moment.

Next consider a machine. It is "born" in some factory, is transported to another, lives out a productive life there, moves into simi-retirement, and finally "dies" on the scrap heap. All ~~of~~ this can again be described by a function whose domain is an interval of $\underline{T}$, and whose range is in $\underline{R} \times \underline{S}$.

An apple is "born" on the branch of a tree, is harvested, moves through the channels of trade to a household, and ends its existence in somebody's stomach. A certain rock was formed ~~about~~ when the Earth was created, and will persist until the Earth is destroyed. And so it goes. ~~The point is that~~ one can think of the world as a concatenation of processes of this type, each representable as a function whose domain is an interval of $\underline{\text{Time}}$, and whose range is in $\underline{R} \times \underline{S}$.

~~There are~~ certain problems [are] connected with this point of view. First, by the law of the conservation of mass, "births" and "deaths" are ~~of course~~ transformations from one form of matter to another. In principle this can be handled by our apparatus: The apple disappears, but the person who has eaten it becomes slightly different in state: less hungry, better nourished, ~~weightier~~ [heavier].

Second, there are ambiguities in the description. Suppose a handle and a blade are combined to make a knife. One possible description of this event is that the handle and blade both cease to exist at this instant, and the knife begins to exist. Another possibility is to have the blade and handle maintain their separate existences, merely being combined from then on in a certain spatial configuration. In this second approach, "knife" is not a resource-type at all, but the name for certain spatial configurations of other resources. (The difficulty here is of a piece with that discussed earlier in the Appendix on Resources of section 3. If the program suggested there could be carried out, it would avoid this problem as well). The solution lies in making conventions as to what is to be considered a resource, as opposed to a spatial configuration of other resources.

A third problem concerns resources that are "continuously" spread over space. The precise meaning of this term will be taken up later, but for the present we may take it to refer to such resources as air, water, soil, wheat, cement and steel, as opposed to people, animals, cars and machines which are more naturally thought of as "discrete" particles.[17] All the examples we have given are of the "discrete" type, and the question arises, can one describe the continuous resources in the same terms? The answer is yes. In fact, one of the great advantages of the measure-theoretic approach is that it can handle discrete, continuous and mixed distributions with equal facility. We shall take this point up later when the approach has been more fully expounded.

We now return to the main line of argument. To repeat, the world is being viewed as a collection of processes, each of which can be represented as a function whose domain is an interval of Time, and whose range is in $\underline{R} \times \underline{S}$.

We need not exclude the possibility that a given process has no birth, so that its existence stretches indefinitely into the past; or that it has no death, so that its existence stretches indefinitely into the future; or neither birth nor death. The following definition formalizes these considerations.

Definition: A history is a function whose domain is a closed T-interval – ~~that is~~, a subset of T (the real line) of one of the following four types: either $\{t|t_1 \leq t \leq t_2\}$, (where $t_1 < t_2$ are real numbers); ~~or~~ $\{t|t \leq t_2\}$; ~~or~~ $\{t|t \geq t_1\}$; or T itself ~~– and which takes values in $R \times S$.~~

(The history of a person – ~~that is~~, a history ~~which~~ takes on only person-types as values in R – may be referred to as a biography.)

Definition: For a given history h, the function ~~which~~ takes on the value s ($\in S$) when h take on the value (r,s) will be called the itinerary of that history.

Definition: The function ~~which~~ takes on the value r ($\in R$) when h takes on the value (r,s) will be called the transmutation-path of that history.

Thus the itinerary traces out the locations occupied by a history in the course of its existence; the transmutation-path traces out the states in the Resources set through which the history passes.

For example, if we take the biography of a person, his itinerary will trace out all his movements, trips, visits, migrations and commuting patterns over his lifetime. His transmutation-path will trace out his progress from infancy to childhood to adulthood to old age, with the accompanying moods, experiences, activities, speech, etc.

We shall denote the itinerary of history h by h_s, and its transmutation-path by h_r. Thus h_r takes values in R, and h_s takes values in S.

Now let Ω be the set of all possible histories — that is, the set of all functions from closed T-intervals to $R \times S$ (not merely those histories realized by an actual "particle"). We now show how the world may be described as a measure space (Ω, Σ, μ) with universe set Ω.

The measure μ has the following intuitive interpretation. For a set of histories $E \in \Sigma$, $\mu(E)$ is the total "mass" flowing through the locations and forms at the times indicated by the various histories of E. This vague characterization will be elucidated in the next few pages.

Along with the measure space (Ω, Σ, μ) we shall consider certain families of functions, all with domain a subset of Ω, and with values in various product spaces built up from R, S, and T. Each such function corresponds to the asking of a question, the answer to which appears as a measure on the space in which its values lie.

Formally, let f be one such function. f takes values in a set A which is typically of the form $R^a \times S^b \times T^c$ for some non-negative integers a, b, c, though it may be more complex. We assume that R, S, and T come supplied with appropriate sigma-fields, Σ_r, Σ_s, Σ_t, respectively, and these determine a product sigma-field on A. Then for any measurable subset $G \subseteq A$, we assume that the set of histories

$$\{h | f(h) \in G\}$$

belongs to Σ — that is, we assume that f is measurable. f then induces the measure μ onto the measurable space A, and this induced measure is, intuitively, the answer to the question embodied in the function f. We now illustrate, beginning with specifics, then generalizing.

Cross-Sectional Measures

Consider the question, "What is the total quantity of water in Lake Erie at Noon, January 26, 1970 (in tons)?" The answer is given by the μ-value of a certain set of histories, E. Specifically, E is the set of histories whose transmutation-paths at the moment Noon, January 26, 1970 lie in the subset of R which is labeled "water", and whose itineraries at that instant are located in the region "Lake Erie." E can be written symbolically as

$$\Big\{h \Big| h(\text{Noon, January 26, 1970}) \in (\text{water} \times \text{Lake Erie})\Big\}. \quad (1)$$

Questions of the general form, "What is the total quantity of resources of types F in region G at time T?" — of which the above is an example — may be called cross-sectional questions. It should be clear that any cross-sectional question (with $F \in \Sigma_r$, $G \in \Sigma_s$,

$t \in T$) has as answer the μ-value of a certain set of histories, ~~namely~~,

$$\{h \mid h(t) \in F \times G\}. \qquad \text{~~(2)~~}$$

~~Let us~~ consider the logic of the situation. The various drops or molecules of water in Lake Erie at the moment in question had a variety of past histories: some fell directly as rain, some flowed in from Lake Huron, some entered as sewage, some as industrial effluent. And they will have a variety of future histories: some evaporating, some flowing out to sea, some entering samples taken by pollution researchers, ~~etc.~~ All of these combinations, and more, will be in the set of histories (1). But cross-sectional questions are not aimed at eliciting this detail; instead they lump together all such histories "from whatever source derived" and "to whatever destiny aimed", and this is just what sets of the form (2) do.

Before giving further examples, ~~let us~~ examine the assumptions behind this whole approach. At first glance, ~~what is involved~~ seems to be a conservation of mass assumption: the same total "quantity of matter" is carried ~~along~~ through time along the paths traced out by the histories, merely changing its form and location (~~that is~~, redistributing itself over $R \times S$). And, indeed, this literal interpretation of "mass" is perfectly adequate for many kinds of histories.

Trouble arises when ~~one~~ considers biographies of persons: ~~It is clear that~~ mass in the literal sanse changes as one advances from infancy to adulthood to corpulent dotage. But one has a certain freedom in choosing measurement units. In the case of resources that come in "natural units" - such as people, cars, or cattle - it is common to measure in terms of "number of entities" rather than in terms of "number of pounds."[18]

Which measurement units to choose? We enunciate the principle: choose measurement units in such a way that the resulting "mass" is conserved as one traces out the path of histories through time. Thus for most social science purposes the "number of persons" measurement approach is the correct one because it gives each person the constant "mass" of 1 over his lifetime.

From now on we shall drop the quotation marks around "mass", it being understood that the appropriate units are being used for the various histories - whether they be pounds, numbers, acres, yards, board-feet, etc. Two points should be noted. First, there is no theoretical objection whatever to adding together measurements using different units for the different components of the sum. So long as the various units are known, the measure μ carries the information without loss. Furthermore, if we switch from one system of measurement units to a completely different system, a simple formula enables one to translate the old measure μ into a new measure μ' in terms of the new units.[19]

Second, it is not clear *a priori* that one *can* define measurement units in such a way that the desired goal of mass conservation is attainable, even approximately. For further discussion, see the Appendix below.

To generalize, consider the function f_t (t being a fixed real number) which assigns to history h the value $h(t)$, which is a point in $R \times S$. Here the domain of f_t consists of those histories which are in existence at instant t. f_t is assumed to be measurable (with respect to Σ restricted to the domain of f_t, and with respect to $\Sigma_r \times \Sigma_s$, the sigma-field of $R \times S$). f_t induces the measure μ onto the space $(R \times S, \Sigma_r \times \Sigma_s)$.

What value is assigned to the measurable rectangle $F \times G$ ($\subseteq R \times S$)? The value which μ assigns to the inverse image

$$\{h \mid f_t(h) \in F \times G\}.$$

But this is the same as the set (2). The measure induced by f_t is the cross-sectional measure giving the distribution of mass over $R \times S$ at time t. This measure provides the answer to any question one wishes to ask concerning the world at time t — that is, to all possible cross-sectional questions. All this information is contained in the original measure μ over the space of possible histories (Ω, Σ) and is extracted from that measure by means of the mapping f_t. Let μ_t be the cross-sectional measure for time t, so that

$$\mu_t(E) = \mu\{h \mid h(t) \in E\},$$

for all $E \in \Sigma_r \times \Sigma_s$.

In general, one will not be interested in the entire realm $R \times S$. A regional geographer who wants to know everything about Austria at time t, for example, will restrict μ_t to $R \times$ Austria.

On the other hand, someone who wants to know everything about steelmaking at t_o, wherever it exists, will restrict μ_{t_o} to $F \times S$, where F is the set of resource-types having to do with steelmaking — ore, coke, slag, blast furnaces, steelworkers, etc. In general one makes both restrictions, narrowing attention to a subset of resources in some region.

This is perhaps the time to bring up the question of practice. Even after restricting one's attention, the resulting measure is a very complicated business. In practice, doesn't one have to simplify drastically in order to say anything at all?

There are three answers to this question. First, one does indeed simplify in practice. The most common method is to aggregate into some simpler sub-sigma-field, usually one generated by a finite partition. The rasult, of course, is still a measure.

The second answer is that it is possible to simplify without aggregating. Practice demands that a description be specifiable by a small number of numerical parameters. Aggregation does this. But it can also be accomplished by having a stock of standard measures available, indexed by a small number of parameters. If the stock is well chosen and versatile, one can find an element which is a good approximation (or "fit") to the actual measure. Examples are the Pearson family of distributions in statistics, and the approximation of functions by polynomials or trigonometric sums. Indeed, the general practice of approximating things by other things in a smaller, simpler family is a universal principle of scientific work; large literatures exist on how to find the best approximation, or test for "goodness of fit".

The third answer refers to the division of labor between practical and theoretical work. Consider numerical calculation. In practice one needs only the rational numbers (or even less say those rationals of the form $N \cdot 10^{-20}$, N an integer). But for theoretical work this would be a crippling restriction. The real numbers are needed even for evolving and justifying practical procedures of numerical calculation itself.

In the same way, even if the only measures ever to be used in a practical way are the aggregations into finite sigma-fields (a premise we do not grant), one would still want to use measure-theory to gain theoretical insight.

We now return to cross-sectional measures. It has been mentioned that "complex" resources may be thought of as <u>spatial configurations</u> of simpler resources. We are now in a position to pin down the concept of "configuration". Consider a certain building at time t, for example, which is a configuration of bricks, wood, plaster, glass, etc. Let E be the region occupied by this building. The configuration which is this building is then simply the cross-sectional measure μ_t restricted to $R \times E$. This restriction tells us how much of every kind of material is present in each part of E, which is just the information we need to describe the building completely. And in general, any "spread-out" entity at a time t may be identified with the cross-sectional measure μ_t restricted to $R \times E$, E being the region occupied by the entity in question.

This takes care, more or less, of a specific entity at a specific time. It is also of interest to define the concept of a type of configuration, not tied down to say any specific region or time-instant. We shall delay giving such a definition until certain further mathematical concepts have been introduced.[25]

Different configurations with the same R-marginal may be referred to as isomers, to borrow a term from chemistry.

So far we have been dealing with facts involving one point in Time. We now go on to facts involving two points in Time, which introduces transformations, transportation and storage.

For example, how many people alive at time t_1 have died by time $t_2 > t_1$? The answer is

$$\mu\left\{h \,\middle|\, h_r(t_1) \in \text{person, and } t_2 \text{ is not in the domain of } h\right\}. \qquad (3)$$

The set of histories in (3) is exactly that called for in the question. (It is assumed here that for histories of this type the measurement units are "numbers of entities", as discussed above).

If instead, t_2 precedes t_1, then (3) gives the number of people who were born between t_2 and t_1, and were still alive at t_1.

How many people migrated from region F_1 at time t_1 to region F_2 at time t_2 $(t_2 > t_1)$? The answer is

$$\mu\left\{h \,\middle|\, h(t_1) \in (\text{person} \times F_1), \text{ and } h(t_2) \in (\text{person} \times F_2)\right\}. \qquad (4)$$

Actually, one should qualify this statement. (4) gives the number of people who were in region F_1 at t_1 and in region F_2 at t_2. Hence, first of all it says nothing about their itineraries within the interval; these may involve all sorts of spatial maneuvers. Secondly, it gives the number of people physically present in these regions rather than being resident in them, change in residence being the usual definition of migration. Residential location could be represented, but it is a more complicated concept than physical location, involving mental states and legal documents.

As a special case of (4) we could have $F_1 = F_2 = F$. Then (4) would count the number of people who stayed in region F throughout the interval, but would also count those who wandered out of the region after time t_1, but returned by time t_2.

How much cotton yarn at time t_1 has been converted into shirts at t_2 ($t_2 > t_1$)? The answer is

$$\mu\left\{h \,\middle|\, h_r(t_1) \in \text{cotton yarn, and } h_r(t_2) \in \text{shirts}\right\}. \qquad (5)$$

(5) gives the mass of the set of histories whose transmutation-path was in the resource-set "cotton yarn" at instant t_1 and in the resource-set "shirts" at instant t_2. Again some qualifications are in order. As above, there is no restriction on what these histories do in the interim period. More serious is the fact that, depending on how histories are defined, (5) may give a "wrong" answer. Recall the discussion of knoves, blades, and handles (p.), where it is pointed out that when a given history is "born"

or "dies" is partly a matter of convention. If things are defined so that cotton yarn ends its existence when converted into shirts, then (5) gives the answer zero. The difficulty hearkens back to the problem of defining the Resources set R in a satisfactory manner.

Instead of considering questions piecemeal, let us set up a measure which answers all such questions systematically. We have already considered the case of a single moment t_0 and the resulting cross-sectional measure on universe set $R \times S$. Now we consider two moments, and get a measure over $(R \times S)^2$. Such measures are called two-timing, (or perhaps double-cross-sectional).

Given two moments, t_1 and t_2 (with $t_1 < t_2$), define the function f_{t_1,t_2} by

$$f_{t_1,t_2}(h) = (h(t_1),\ h(t_2)). \tag{6}$$

The domain of f_{t_1,t_2} is the subset of histories which are in existence at both times t_1 and t_2. It is assumed that f_{t_1,t_2} is measurable. Hence from μ it induces a measure μ_{t_1,t_2} on the range space $(R \times S \times R \times S)$.

The intuitive meaning of μ_{t_1,t_2} is as follows. Let E and F be measurable subsets of $R \times S$. Then $\mu_{t_1,t_2}(E \times F)$ is the total mass of all histories having a value in E at moment t_1, and in F at moment t_2.

Again, we may let the sets E and F themselves be rectangles in $R \times S$: Let $G_1, G_2 \in \Sigma_r$ and $H_1, H_2 \in \Sigma_s$. Then

$$\mu_{t_1,t_2}[(G_1 \times H_1) \times (G_2 \times H_2)]$$

is the mass embodied in the histories which are in resource set G_1 and region H_1 at time t_1, and move to resource set G_2 and region H_2 at time t_2.

This measure gives no information concerning histories which are "born" or "die" between t_1 and t_2. To answer such questions systematically one can proceed as follows. (Details concerning measurability are omitted). We add an artificial point z_o — signifying non-existence — to the set $R \times S$. f_{t_1,t_2} is again defined as in (6), but its domain is now all of Ω; if history h is not in existence at time t_i then $h(t_i)$ is to be understood as z_o. This extended function induces a measure μ_{t_1,t_2} onto the space $[(R \times S) \cup \{z_o\}]^2$. For example,

$$\mu_{t_1,t_2}[(\text{person} \times S) \times \{z_o\}]$$

would be the same as (3) — the total number of persons alive at time t_1 who have died by t_2.

Having gone from one to two time-points, it is simple to go to three, to a finite, or even to a countable number of time points. For example, choose a measurable set $E_t \subseteq [(R \times S) \cup \{z_o\}]$

for each <u>integer</u> t ($\pm$ 1, $\pm$ 2,...). The measure of

$$\{h \mid h(t) \in E_t,\ t = 0, \pm 1, \pm 2, \ldots\}$$

gives the mass of all histories passing through each of the sets E_t at the respective integer times. One could even do this for all the <u>rational</u> instants t, since these are countable — as good a monitoring system as one could hope for.

Production and Consumption

A broad category of questions concerns births or production over time, such as "How much corn was grown in Iowa in 1948?" or "How many people were born in New York in 1934?"

To give a general method for drawing such descriptions out of the measure space of histories (Ω, Σ, μ), we first restrict Ω to the subset Ω_- consisting of all histories which have a date of birth — that is, do not exist indefinitely for back into the past. Then define the function $f: \Omega_- \to R \times S \times T$ by:

$$f(h) = (h(t_1), t_1) \tag{7}$$

where t_1 is the moment of birth of the history h. That is, f assigns a pair, the second component of which is the moment of birth (= the earliest time at which h takes a value in $R \times S$), and the first is the value which h takes at that time.

In terms of f the amount of corn grown in Iowa in 1948 is the measure of the set of histories

$$\{h \mid f(h) \in \text{corn} \times \text{Iowa} \times 1948\}, \qquad (8)$$

since (8) is precisely the set of histories which are "born" in the time interval 1948, whose transmutation-path starts in the resource-set "corn", and whose itinerary starts in the region "Iowa".

Assuming f to be measurable, it induces the measure μ (restricted to Ω_-) onto $(R \times S \times T, \Sigma_r \times \Sigma_s \times \Sigma_t)$. Let us call this induced measure λ_1. On rectangles λ_1 can be given a simple intuitive interpretation: Let $E \in \Sigma_r$, $F \in \Sigma_s$, $G \in \Sigma_t$; then $\lambda_1(E \times F \times G)$ = total mass of all histories starting at some instant in period G in region F in resource-set E. Thus λ_1 gives the distribution of "births" or "production" over Resources, Space and Time.

By an argument exactly parallel to the one just given we can describe the distribution of "deaths" or "consumption". Omitting details, we restrict Ω to those histories having an end in Time, then take a function g having as its value the pair consisting of the date of death and the point in $R \times S$ occupied by the history at that moment. Assuming g is measurable, it induces a measure λ_2 onto $(R \times S \times T, \Sigma_r \times \Sigma_s \times \Sigma_t)$. The interpretation of λ_2 on rectangles is:

$\lambda_2(E \times F \times G)$ = total mass of all histories ending at some instant in period G in region F in resource-set E.

Finally, we consider the joint pattern of production and consumption. First, restrict Ω to the set of histories having both a beginning and an end; call this Ω_o. Now define the function

$$k: \Omega_o \to (R \times S \times T \times R \times S \times T) \; (= (R \times S \times T)^2, \text{ for short})$$

given by $k(h) = (f(h), g(h))$,

where f is defined by (7), and g is the similar function defined above. That is, $k(h)$ is a quadruple giving the point in $R \times S$ at which h starts, the time it starts, the point in $R \times S$ at which it ends, and the time it ends.

On $(R \times S \times T)^2$ we take the sixtuple product field $(\Sigma_r \times \Sigma_s \times \Sigma_t)^2$. Since f and g are measurable, so is k. Let ν be the measure it induces on $((R \times S \times T)^2, (\Sigma_r \times \Sigma_s \times \Sigma_t)^2)$. On rectangles the interpretation of ν is as follows:

Let $E_1, E_2 \in \Sigma_r$; $F_1, F_2 \in \Sigma_s$; $G_1, G_2 \in \Sigma_t$; then $\nu(E_1 \times F_1 \times G_1 \times E_2 \times F_2 \times G_2)$ = total mass of all histories starting at some instant in period G_1 in region F_1 in resource-set E_1, and ending at some instant in period G_2 in region F_2 in resource-set E_2.

If we think of $(R \times S \times T)^2$ as the product of $(R \times S \times T)$ by itself, then λ_1 and λ_2 are precisely the left and right marginal measures of ν, respectively.

By this time the main lines of development of our descriptive program should now be clear. "Extensive magnitudes" in general may be represented as measures, and a large variety of these may be derived from one underlying measure μ on the space of histories. We could, in fact, extend this section indefinitely, systematically deriving more and more complex varieties of data from the underlying measure space (Ω, Σ, μ); but this would begin to strain the reader's patience.

(As an exercise, the reader is invited to puzzle out how the fishing example that begins this chapter may be derived from (Ω, Σ, μ). This is more complex than our previous cases, because, e.g., codfish are not produced in the port of Boston, but arrive there from elsewhere. The solution — as a restriction of a measure over $\underline{R} \times \underline{S} \times \underline{T}$ — involves counting each history the number of times it enters a given subset of $\underline{R} \times \underline{S}$.)

One final comment on the scope of this program. Our examples have been drawn exclusively from statistical data — that is, the kind of data that appear in tabular numerical form in census reports, etc. These data have a certain precision which makes them easy to discuss. However, since our model deals with the redistribution of matter in the most general sense of the term, both in location and in form (= resource-state), in principle it should be able to handle "literary" data as well — history, travel, biography, belles lettres, etc. "He swept her up in a passionate embrace" could be translated into the language of measures; the only conceptual difficulty lies in the vagueness of the description.

Appendix on Histories

One disconcerting feature of our model is the extreme generality of the concept of "history". Between birth and death any function with values in $\underline{R} \times \underline{S}$, however erratic, is an admissible history.

This in itself is not disqualifying. If, in the real world, frogs do not turn into princes, and Dr. Jekyll does not become Mr. Hyde, this is indicated by assigning the measure zero to the appropriate set of histories. But difficulties remain.

Trouble arises from the diversity of measurement units. The more a given history wanders over the set of Resources, the harder it becomes to assign units in such a way that "mass" is preserved over time. This suggests the following kind of modification — (or rather, restriction) — on the set of histories Ω. The set of Resources is given a partition, $\mathcal{R}$, into measurable subsets, such that the elements of any set $E \in \mathcal{R}$ are "similar" to each other in some sense. In particular, they are similar in the sense that the same kinds of measurement units are applicable to all the elements of any given set of the partition. (Thus one would not put into the same set resource types which come in "natural units" and resource types which lend themselves to measurement by weight). ~~This is all rather vague, of course, and is offered merely as one line of approach.~~

Having set up the partition, $\mathcal{R}$, of R into fairly "homogeneous" subsets, we now admit only those histories whose transmutation paths stay entirely within some one set $E \in \mathcal{R}$. This restriction on the set of histories alleviates the measurement unit problem. Each $E \in \mathcal{R}$ may now be tagged with its "natural" unit, whether it be (pounds, number of entities, acres, etc). In setting up a sigma-field on the restricted Ω, one may begin by taking the set of all histories with R-values in set E to be measurable for each $E \in R$. (In contrast to R, Space requires no such "breaking up". Because of its homogeneous nature, the wandering of itineraries over S creates no measurement unit problems.)

The restriction just discussed is a kind of "boundedness" constraint, limiting the "distance" over which any transmutation-path is allowed to wander in R. A different kind of restriction also suggests itself — one prohibiting "discontinuous jumps". (The quotation marks are used in this paragraph because ~~so far~~ we have not yet defined any structure on R or S ~~which~~ that would give meaning to them.)[22]

Without going into ~~any~~ details ~~at present~~ here, suppose the concept of continuity for histories has been defined, and in such a way that the maxim, natura non facit saltum, is valid.[23] This ~~fact again~~ does not by itself disqualify our original scheme. It just means that measurable sets of discontinuous histories ~~get~~ are assigned the value zero. However, in this case there may be some advantage to restricting our original Ω to the simpler subset of continuous histories.

Finally, we mention the measurability problem on the space of histories. ~~As we have discussed~~ above, this is the problem of identifying which sets of histories correspond to observations that might be made, at least conceptually. The criterion of "conceptual observability" is itself vague; but even if it were pinned down one would still have to classify systematically the possible kinds of data and find the subsets of Ω corresponding to each.

2.6 Measure Theory, III

~~We return now to the exposition of measure theory.~~ This section differs in style from 2.1 and 2.4 in two respects. First, the ratio of theorems to definitions is higher: We ~~shall be~~ more concerned with stating the results of the theory, and less with merely outlining the concepts ~~of the theory~~. Second we ~~shall~~ give illustrations not only from pure mathematics, but also from the applied concepts we have been building ~~up~~ (R, S, T, histories, etc.) No confusion should result from this mixture. As before, we ~~shall~~ omit proofs unless they are very short or instructive, or not readily available.

Finite and Sigma-Finite Measures

A first destinction is between finite and infinite measures, the latter being those ~~that~~ take on the value ∞ at least once. Since a measure attains its maximum value on the universe set A, a measure μ is finite iff μ(A) is finite, and infinite iff μ(A) = ∞.

Definitions: Consider any function f whose range is in the extended real numbers. f is finite above iff it never takes on the value +∞, finite below iff it never takes on the value −∞, finite or real-valued iff it is finite both above and below. On the other hand, f is bounded above (bounded below) iff there is a real number L such that $f(x) \leq L$ ($f(x) \geq L$) for all x in the domain of f. f is bounded iff it is bounded both above and below. Equivalently, f is bounded iff there is a real number L such that $-L \leq f(x) \leq L$ for all x in the domain of f, which can also be expressed by writing $|f(x)| \leq L$, all x, vertical bars indicating the absolute value of a number ($|a| = a$ if $a \geq 0$; $|a| = -a$ if $a < 0$).

Note that f bounded above, below, or both implies f finite above, below, or both, respectively, but ~~that~~ the converse is not necessarily true. For example, the identity function $f(x) = x$ on the real line is finite but not bounded.

However, for measures, the properties "finite", "finite above", "bounded", and "bounded above" are all equivalent. To see this, first note that measures are automatically bounded below, since they are non-negative; second, if μ is finite, the real number $\mu(A)$ provides an upper bound: $\mu(E) \leq \mu(A)$ for all measurable E. ~~From now on,~~ the terms "bounded measure" and "finite measure" will be used interchangeably.

A related very important concept is ~~sigma~~-finiteness.

Definition: Let (A, Σ, μ) be a measure space. μ is ~~sigma~~-finite iff there is a countable partition, $\mathcal{G}$, of A into measurable sets, such that, for each $G \in \mathcal{G}$, the restriction of μ to G is finite.

Examples:

~~(i)~~ Any finite measure is ~~sigma~~-finite. (Proof: let the partition $\mathcal{G}$ have as its only member the universe set A itself).

~~(ii)~~ Consider Lebesgue measure on the real line. This is certainly not finite, since it assigns the value ∞ to the entire line. But it is ~~sigma~~-finite. (Proof: take the countable measurable partition consisting of the sets $\{x|n \leq x < n + 1\}$, where n runs through the integers 0, ± 1, ± 2, ...; μ is finite on each piece — in fact $\mu\{x|n \leq x < n + 1\} = 1$, all n).

(iii) Any measure which assigns the value ∞ to some singleton set $\{x\}$ is not sigma-finite. (Proof: for any partition $\mathcal{G}$, if $x \in G \in \mathcal{G}$, then the restriction of μ to G remains infinite.)

(iv) As a less trivial example of a non-sigma-finite measure, let A be the real line, Σ any sigma-field on A, and μ the enumeration measure ($\mu(E)$ = number of points in E). (Proof: let $\mathcal{G}$ be any countable packing of measurable sets such that $\mu(G)$ is finite for all $G \in \mathcal{G}$; then each G is finite, so $\cup\mathcal{G}$ is a countable set; since A is uncountable, $\mathcal{G}$ cannot be a partition).

The importance of sigma-finite measures stems from two facts: They have many useful properties not shared by measures in general, and most measures which come up, even in theoretical investigations, are sigma-finite.

Atomic and Non-Atomic Measures

We have already mentioned informally the distinction between resources which are typically distributed "discretely" over Space and those which are distributed "continuously", and we now want to define these concepts rigorously and abstractly. Actually, it turns out that there are two entirely different concepts explicating the notion of "continuous distribution", one of them involving a single measure, the other a certain relation between two measures. We give the first one now.

Definition: A measure μ is *non-atomic* iff, for any measurable set E for which $\mu(E) > 0$, there is a pair of measurable sets F, G such that $F \cap G = \emptyset$, $F \cup G = E$, $\mu(F) > 0$ and $\mu(G) > 0$.

Briefly, μ is non-atomic iff any set of positive measure can be split into two pieces, each of positive measure.

Examples:

(i) It may be shown that Lebesgue measure on the real line is non-atomic.

(ii) Let $\mu(\{x\}) > 0$ for some measurable singleton set $\{x\}$; then μ cannot be non-atomic (since $\{x\}$ cannot be split).

(The converse of this statement is not true: There are measures for which $\mu(\{x\}) = 0$, all x, yet which are *not* non-atomic).

Definition: Given a measure space (A,Σ,μ), a set $E \in \Sigma$ is called an *atom* for μ iff $\mu(E) > 0$, and, however E is split into two measurable sets F, G ($F \cap G = \emptyset$, $F \cup G = E$), either $\mu(F) = 0$ or $\mu(G) = 0$. (Thus a measure is non-atomic iff Σ contains no atoms.)

At the other extreme we have

Definition: Let (A,Σ,μ) be a measure space. μ is an *atomic measure* iff A itself is an atom.

That is, $\mu(A) > 0$, and, for any $E \in \Sigma$, either $\mu(E) = 0$ or $\mu(A \setminus E) = 0$. If μ is finite, then μ is atomic iff its range consists of exactly two values, 0 and $\mu(A)$. (This follows at once from the equation $\mu(A) = \mu(E) + \mu(A \setminus E)$ and the definition).

Definition: μ is simply-concentrated iff there is a point $a_o \in A$ having the property:

$$\mu(E) = 0 \text{ if } a_o \notin E, \quad \mu(E) = \mu(A) > 0 \text{ if } a_o \in E, \qquad (2.6.1)$$

for all measurable E.

A simply concentrated measure is atomic; although not all atomic measures are simply concentrated, the latter is the most important kind found in practice. Choosing an arbitrary point a_o, an arbitrary positive number for $\mu(A)$, and μ according to (1) gives a simple recipe for constructing atomic measures (the point a_o need not be unique, in general).

Definition: μ is a sigma-atomic measure iff there is a countable measurable partition, G, of A, such that G is an atom for all $G \in G$. Otherwise expressed, μ is sigma-atomic iff the universe set A can be split into a countable number of measurable pieces, such that μ restricted to each piece is an atomic measure.

Theorem: ("atomic decomposition theorem") Given measure space (A, Σ, μ), with μ sigma-finite. Then there is a set $E \in \Sigma$ such that

(i) μ restricted to E is non-atomic, and

(ii) μ restricted to $A \setminus E$ is sigma-atomic.

If E' is another set satisfying this theorem, then $\mu(E \setminus E') = 0$ and $\mu(E' \setminus E) = 0$: E is "almost" unique. Furthermore, the atoms in any two such decompositions may be paired off such that a similar relation holds between each pair of atoms.

This is the first of several basic decomposition theorems whose aim is to represent measures as built up in one way or another from simpler measures.

We illustrate with a real-world example. Consider the rural-urban distribution of population over the surface of the Earth. It is a useful approximation to think of the urban population as being concentrated in cities each located at a single point of the Earth's surface — (say at $a_1, a_2, \ldots$) — while the rural population is "smeared" over the surface. If Σ for this example is the Borel field — so that all singleton sets $\{a_i\} \in \Sigma$ — then an atomic decomposition is clearly given by $A \backslash E = \{a_1, a_2, \ldots\}$. That is, on this "urban set" population distribution is sigma-atomic (each singleton set $\{a_i\}$ being an atom), and on the complementary "rural set" it is non-atomic.[24]

Having decomposed μ atomically, one is then in a position to take advantage of the special properties of each part. For non-atomic measures the following property is very useful.

Theorem: Given (A,Σ,μ), with μ a non-atomic measure. Let $\mu(E)$ be finite. Then for any real number x such that $0 \leq x \leq \mu(E)$, there is a set $F \in \Sigma$ such that $\mu(F) = x$.

If μ is finite and non-atomic, this means that it takes on every single real value in the interval from 0 to $\mu(A)$. Suppose next that μ is infinite but sigma-finite. It follows easily from the definition that for any real number x, μ takes on a real value greater than x. Combining this observation with the theorem just stated, we conclude: An infinite, sigma-finite, non-atomic measure takes on all positive real numbers as values.

Integration

We start with a measurable space (A,Σ). The integral will be a certain function which assigns an extended real number to every pair consisting of (i) a measure μ on (A,Σ) and (ii) a measurable function f with domain A, and with values in the non-negative extended real numbers.[25] Our notation for the integral is

$$\int_A f\, d\mu \quad \text{or} \quad \int_A f(x)\, \mu(dx). \tag{2}$$

We start with a certain special kind of function f.

Definition: Function f (with domain A) is simple iff it is measurable, real-valued, non-negative, and takes on only a finite number of values.

As an example, the constant function $f(x) = c$ (where $\infty > c \geq 0$) is simple. Another example, which merits a definition of its own, is the following.

Definition: The indicator function of set E (notation I_E) is given by $I_E(a) = 1$ if $a \in E$; $I_E(a) = 0$ if $a \notin E$.

The indicator function of any measurable set E is simple.

Now let f be simple, and let $\{x_1, \ldots, x_n\}$ be its range. Since f is measurable, each set $\{a|f(a) = x_i\}$ is measurable, and the collection of these, for $i = 1, \ldots, n$, constitutes a finite partition of A. We now define $\int_A f\, d\mu$ to equal

$$x_1\mu\{a|f(a) = x_1\} + \ldots + x_n\mu\{a|f(a) = x_n\}. \qquad (2.6.3)$$

(In evaluating (3), recall the rules of arithmetic in the extended real number system. In particular, $x \cdot \infty = \infty$ if $x > 0$, and $0 \cdot \infty = 0$).

Examples: (i) For the constant function $f(x) = c$, (3) consists of just one term, $c \cdot \mu(A)$. (ii) For indicator functions,

$$\int_A I_E\, d\mu = 1 \cdot \mu(E) = \mu(E).$$

We now define the integral in general in terms of its value for simple functions. We use the notation $f \geq g$ for two functions on A to indicate that $f(a) \geq g(a)$ for all $a \in A$. Also, "sup" abbreviates "supremum".

Definition: Given measure space (A,Σ,μ), and non-negative measurable function f on A,

$$\int_A f\, d\mu = \sup\left\{\int_A g\, d\mu \;\middle|\; g \leq f,\ g \text{ simple}\right\}. \qquad (2.6.4)$$

That is, we consider the set of all simple functions bounded above by f; for each of these we form its integral, and the integral of f is defined as the supremum of the resulting set of extended real numbers.[26]

Note that (4) is not circular, since the integral of simple function has already been defined by (3). (4) is also consistent, in the sense that, if f itself is a simple function, then (4) gives the same answer as (3).

A useful extension of this definition is to integration over a measurable subset, E, of A. This is denoted $\int_E f\, d\mu$, and is simply the ordinary integral (4) when f and μ are both restricted to E. (For $E = \emptyset$, we set it equal to zero.) This may also be written as an integral over A. In fact,

$$\int_E f\, d\mu = \int_A I_E \cdot f\, d\mu$$

for all $E \in \Sigma$. (The function being integrated on the right is the product of f and the indicator function of E, so that it coincides with f for points of E, and is identically zero off E).

Let us compare this with the ordinary Riemann integral. Let f be real-valued, continuous and non-negative on the closed interval $\{x | a \leq x \leq b\}$ (a, b real numbers). Then

$$\int_a^b f(x)dx = \int_{\{x|a\leq x\leq b\}} f\, d\mu. \qquad (5)$$

Here the left-hand expression is the Riemann integral in its usual notation, μ on the right is Lebesgue measure, and (5) shows how to translate the Riemann integral into the form (2). (5) is valid for any function f having a Riemann integral. (Actually, one needs a slightly richer sigma-field -- the "Lebesgue completion" of the Borel field -- for (5) to be valid for any such f. This concept is unimportant for our purposes and we pass over it.)

The integral (2)-(4) constitutes a triple generalization of the Riemann integral. First, the class of functions f possessing an integral is broadened. Second, the integral is defined not only for Lebesgue measure but for measures in general. Third, the integral is defined for any abstract measurable space, not just the real line.

In view of this enormous generality the following theorem is surprising, because it shows that the general integral can be expressed in terms of the Riemann integral -- in fact, as the Riemann integral of a monotone non-increasing function.

Theorem: Let (A,Σ,μ) be a measure space, and f a measurable non-negative function on A. Then

$$\int_A f\,d\mu = \int_0^\infty \mu\{a|f(a) > t\}dt = \int_0^\infty \mu\{a|f(a) \geq t\}dt. \qquad (6)$$

(Here the middle and right expressions are improper Riemann integrals defined by the usual limiting processes. Since the integrands are monotonic there is no problem of existence, though $+\infty$ is a possible value.) (6) is proved by comparing the Riemann sums approximately the middle and right integrals with the integrals of simple functions approximating the left expression. The middle or right-hand form in (6) will be referred to as the Young integral.[27]

To illustrate (6), take the constant function $f(x) = c$ $(c > 0)$. $\{a | f(a) > t\} = A$ if $t < c$, and $= \emptyset$ if $t \geq c$; hence the middle integrand equals $\mu(A)$ up to $t = c$, and equals 0 beyond that point; the right-hand integrand is identical except at the single point $t = c$; hence both ~~of these~~ integrals equal $c \cdot \mu(A)$, which we have already verified to be the value of $\int_A c\, d\mu$.

We ~~shall~~ list some standard properties of the integral. f and g are assumed to be measurable non-negative extended real valued functions on A.

$$\int_A f\, d\mu \geq 0. \qquad (7)$$

If $f > 0$ (~~that is~~, $f(a) > 0$ for all a), and $\mu(A) > 0$, then

$$\int_A f\, d\mu > 0. \qquad (8)$$

If f and μ are both bounded, then $\int_A f\, d\mu$ is finite. (9)

If c is a positive number, then

$$c\int_A f\, d\mu = \int_A cf\, d\mu. \qquad (10)$$

$$\int_A f\, d\mu + \int_A g\, d\mu = \int_A (f + g)\, d\mu. \qquad (11)$$

Indefinite Integrals

We have defined $\int_E f\, d\mu$, where E is a measurable set. Now, for fixed f and μ (where these are defined as above), consider the function ν, with domain Σ, which is given by

$$\nu(E) = \int_E f\, d\mu.$$

Definition: ν is known as the indefinite integral of f with respect to μ. We ~~shall~~ use the notation $\int f\, d\mu$ for the indefinite integral.

Theorem: $\int f\, d\mu$ is a measure. If f and μ are both bounded, then $\int f\, d\mu$ is bounded. If f is finite, and μ is sigma-finite, then $\int f\, d\mu$ is sigma-finite.

The first statement may be proved by aid of the monotone convergence theorem, which we come to later. We ~~shall~~ prove the last two statements. If f and μ are both bounded, the boundedness of $\int f\, d\mu$ is immediate from (9). Let f be finite and μ sigma-finite, and let $\{G_m\}$, $m = 1, 2, \ldots,$ be a measurable partition of A such that $\mu(G_m)$ is finite for all m. Let

$$E_{mn} = \{a \mid a \in G_m \text{ and } n \le f(a) < n + 1\},$$

where $m = 1, 2, \ldots,$ and $n = 0, 1, 2, \ldots$. The class of sets

E_{mn} is a countable measurable partition of A, and f and μ are both bounded on each piece. Hence by (9), $\int_{E_{mn}} f\,d\mu$ is finite, all m, n, so that $\int f\,d\mu$ is sigma-finite. ∎

One application of indefinite integrals is to the problem of change in measurement units, which we left hanging (see p.). Let (A,Σ,μ) be a measure space representing some real-world data. Measurement units need not be homogeneous: they may be "acres" in one portion of A, "pounds" in another, "numbers of entities" in a third, etc.[28] Now suppose the measurement units are changed in some arbitrary manner. The same data will now be represented by a new measure, ν, in terms of the new measurement units. For example, if everything were previously measured in kilograms, and we convert to grams, obviously μ gets blown up by a factor of 1000:

$$\nu(E) = 1000\,\mu(E), \text{ all } E \in \Sigma. \qquad (12)$$

But what is the general relation between μ and ν?

The change in measurement units can be represented by a function f on A: $f(a)$ = number of new units equivalent to one old unit at point a. f is obviously real-valued and positive. The only other restriction we impose is that it be measurable. The relation between μ, ν and f is then

$$\nu = \int f\,d\mu. \qquad (13)$$

As an example, take the conversion above of kilograms into grams. In this case $f(a) = 1000$ for all $a \in A$. Checking the formula for the integral of a constant, we see that (13) does, indeed, reduce to (12) in this case.

Let μ be a measure, and f, g two non-negative measurable functions on (A,Σ). Since the indefinite integral $\int g\, d\mu$ is a measure, one can integrate f with respect to *it*.

Theorem: Let (A,Σ,μ), f, and g be as stated. Then

$$\int f\, d\left(\int g\, d\mu\right) = \int fg\, d\mu. \tag{14}$$

That is, the indefinite integral of f with respect to the measure $\int g\, d\mu$ is the same as the indefinite integral of fg with respect to μ.

As an illustration take the measurement-unit transformation discussed above. Suppose one changes measurement units according to function g, and then changes them again according to function f. The composites result is the left-hand expression in (14), and this theorem states that this yields the same transformation as a single change in units represented by $h(a) = f(a) \cdot g(a)$.

In particular, if we simply invert the previous change, then $f = 1/g$, and, by (14),

$$\int \frac{1}{g}\, d\left(\int g\, d\mu\right) = \int 1\, d\mu = \mu,$$

as it should.

Densities

We pull together two ~~threads~~ in this section. ~~In the~~ first ~~place,~~ we have mentioned that the intuitive concept of "continuous distribution" has two quite distinct explications. One ~~of these~~ is the concept of "non-atomic" measure, which we have discussed. The second, "absolute continuity", will be taken up here.

Second, we mentioned at the ~~very~~ beginning of this chapter that certain kinds of data — such as prices or population densities — which could not themselves be represented as measures, could be derived from measures in a certain way. The same circle of ideas serves to accomplish this.

Definition: Let μ, ν be two measures on the same measurable space $(\underline{A},\Sigma)$. ν is absolutely continuous with respect to μ iff, whenever $\mu(\underline{E}) = 0$, then $\nu(\underline{E}) = 0$. The notation for this state of affairs is: $\nu \ll \mu$.

It follows at once that absolute continuity is transitive: if λ, μ, ν are three measures on $(\underline{A},\Sigma)$ such that $\nu \ll \mu$ and $\mu \ll \lambda$, then $\nu \ll \lambda$. Also, of course, $\mu \ll \mu$.

Theorem: Let ν be the indefinite integral $\int f\, d\mu$. Then $\nu \ll \mu$.

Proof: Suppose $\mu(\underline{E}) = 0$ for $\underline{E} \in \Sigma$, so that μ restricted to $\underline{E}$ is identically zero. If $\underline{g}$ is a simple function, it follows from (3) that

$$\int_E g \, d\mu = 0.$$

Then, from (4), this must be true for any non-negative measurable function g, in particular for f. Hence $\nu(E) = 0$. ///

The basic result concerning absolute continuity is that, under slight restriction, the converse of this statement is true:

Theorem: (Radon-Nikodym theorem). Let μ, ν be two measures over (A,Σ), such that μ is sigma-finite, and $\nu \ll \mu$. Then there exists a non-negative measurable function f such that

$$\nu = \int f \, d\mu. \tag{15}$$

(f is known as the density, or Radon-Nikodym derivative, of ν with respect to μ, and is sometimes written $d\nu/d\mu$.)

To make a statement concerning the extent to which the density f is uniquely determined by ν and μ, we need the following concepts.

Definition: Let (A,Σ,μ) be a measure space; a property P is said to hold μ-almost everywhere, or for μ-almost all points, iff there is a set $E \in \Sigma$ such that $\mu(E) = 0$, and P is true for all points of $A \setminus E$. Another way of expressing the same thing is in terms of null sets. A set F is μ-null iff there is a set $E \in \Sigma$

such that $F \subseteq E$ and $\mu(E) = 0$. Then P holds μ-almost everywhere iff the set $\{a|P \text{ is not true for } a\}$ is μ-null. In all of this, if μ is understood it may be omitted; thus one says simply "almost everywhere", etc.

Definition: Two functions f, g are μ-equivalent (or μ-almost identical) iff the set $\{a|f(a) \neq g(a)\}$ is a μ-null set.

Theorem: Let f, g be two non-negative measurable functions, and μ a sigma-finite measure, on (A, Σ). Then for the indefinite integrals we have

$$\int f\, d\mu = \int g\, d\mu$$

iff f and g are μ-equivalent.

This answers the uniqueness question concerning the Radon-Nikodym derivative. In that theorem only μ is required to be sigma-finite. If ν is also sigma-finite we can state the stronger conclusion that there is a finite density f satisfying (15).

Let us give some possible real-world examples, and in fact let us compare all this with the intuitive concept of "density". Take 3-dimensional space with the Borel field. Just as the ordinary concept of length extends to Lebesgue measure on the real line, the ordinary concept of volume extends to a measure known as 3-dimensional Lebesgue measure in 3-space.[29] For simplicity, let us continue to refer to this extension as "volume" and denote it by μ. Let ν be the mass distribution of some resource-type over space. The average density of this resource in a region E of positive volume is given by $\nu(E)/\mu(E)$. (Average density is not defined if $\mu(E) = 0$). Average density is thus a set function, whose domain is a certain subclass of the Borel field Σ. This is rather unwieldy, and one would like to go from average density to density at a point. (A rough analogy is the process of going from average slopes to the more useful derivatives).

The Radon-Nikodym theorem pins down these vague notions. We assume that ν is absolutely continuous with respect to volume, μ. (i.e., for any region E, if E has no volume, E has no resource content). Since μ is sigma-finite, it follows that there exists a point-function f satisfying (15), and this is exactly the property one would want a point-density to have.

Now the foregoing analysis did not depend in any way on the particular natures of the two measures involved. This raises the possibility of thinking of a great many other types of data as being "densities" derived in this way from two measures. We give several examples:

(1). Let μ be population distribution over the surface of the Earth, by place of residence (measurement unit = number of people). Let ν be the distribution of total income, again attributed by residence. We see that $\nu \ll \mu$, since no income accrues to unpopulated regions. The "density" $d\nu/d\mu$ in this case is simply per-capita income.

(ii) Let μ be the distribution of economic commodities over Space, measured in mass units, perhaps quite heterogeneous. Let ν be the same distribution measured in value or wealth terms (unit: dollars). Again ν << μ. The "density" dν/dμ in this case may be interpreted as prices. In somewhat more detail: The universe set is a subset of R × S, and the density p(r,s) is then the price of resource-type r at location s. (The units in which p(r,s) is measured will be: dollars per acre, or gram, or litre, etc., corresponding to whatever units μ was measured in at point (r,s).)

(iii) Consider the concept of the "quality" of resources: Gold has higher quality than brass, etc. One explication of this somewhat elusive concept is to define quality as the ratio of value to weight. Thus if we let μ be the distribution of resources by weight and ν their distribution by value, "quality" comes out as the density dν/dμ.

(iv) This and the next example show that index numbers may be construed as densities. We are given two terms t_0, t_1, with $t_0 < t_1$. The universe set A is some appropriate subset of R or R × S. There are price systems, p_0, p_1 at times t_0, t_1 respectively. Formally, p_0 and p_1 are measurable positive functions on the universe set. (Prices may be given directly, or may themselves be derived as densities, as under (ii)). Also there are two quantity measures, μ_0, μ_1, referred to the respective times. These may represent stocks, or production, or consumption, or exports, etc. We suppose that each of these measures is absolutely continuous with respect to the other; that is, $\mu_0(E) = 0$ iff $\mu_1(E) = 0$, for all measurable sets $E \subseteq A$.

The Laspeyres price index for measurable set $E \subseteq A$ is now defined as

$$\frac{\int_E p_1 \, d\mu_0}{\int_E p_0 \, d\mu_0}. \qquad (16)$$

(The Paasche price index substitutes μ_1 for μ_0 in (16).) As E varies, the numerator and denominator of (16) define indefinite integrals, and the price index comes out as an average density of these. The point-density, or Radon-Nikodym derivative of $\int p_1 \, d\mu_0$ with respect to $\int p_0 \, d\mu_0$, is simply $f(a) = p_1(a)/p_0(a)$, since, by (14),

$$\int f \, d\left(\int p_0 \, d\mu_0\right) = \int f \cdot p_0 \, d\mu_0 = \int p_1 \, d\mu_0. \qquad (17)$$

(v) The Laspeyres quantity index for measurable E is defined as

$$\frac{\int_E p_0 \, d\mu_1}{\int_E p_0 \, d\mu_0}. \qquad (17)$$

(The Paasche quantity index substitutes p_1 for p_0 in (17).) Again this is the average density derived from two indefinite integrals. The point density of $\int p_0 \, d\mu_1$ with respect to

$\int p_\theta \, d\mu_\theta$ is simply $d\mu_1/d\mu_\theta$, since for $f = d\mu_1/d\mu_\theta$ we obtain, by (14),

$$\int p_\theta \, d\mu_1 = \int p_\theta \, d\left(\int f \, d\mu_\theta\right) = \int p_\theta f \, d\mu_\theta = \int f \, d\left(\int p_\theta \, d\mu_\theta\right).$$

These examples should illustrate the variety of data which can be brought under the rubric "density". In fact, examination of statistical compilations would show that, of the data which cannot be represented directly as measures, the great bulk can be represented as densities with respect to some pair of measures.[31]

In our examples above we have presented the pair of measures first and derived the density from them. It should be noted, however, that in some cases the density is more readily observable than one of the measures (in which case the measure may be constructed as an indefinite integral).

Consider, for example, the standard exercise in capital theory of converting from current to discounted dollars. Let μ be a measure with universe set *Time*, having the interpretation: $\mu(E)$ = value in current dollars of that portion of an income-stream arriving in time-period E. (The use of measure language enables one to cover the cases of lump-sum accruals, continuous accruals, and mixtures of the two all in a single notation). Assuming for simplicity a constant discount rate, i, and discounting to moment t_0, the income stream expressed in discounted dollars is simply the indefinite integral

$$\int e^{-i(t-t_o)}\ \mu(dt). \qquad (18)$$

Here the density $f(t) = e^{-i(t-t_o)}$ is more or less directly observable, and the discounted income stream is constructed from it.[32]

Again, consider prices. One can observe "list prices", if these exist. Or, one can take the ratio of money passing in one direction to goods passing in the opposite direction. The first gives a direct observation of a density (perhaps a misleading one if there are trade discounts, etc.) The second derives price as an average density of two measures.

Before dropping this topic, let us consider the concept of "uniformity". One speaks, especially in spatial economics, of "uniform" population distribution, "uniform" resources, "uniform" planes, etc. A moment's reflection indicates that what these terms are expressing is the proportionality of the measure in question to some other implicit measure, usually surface area. Thus if μ is a real measure and ν is population distribution, then the ascertion is that there is a number c such that $\nu(E) = c\mu(E)$ for all measurable sets E $(0 < c < \infty)$. This in turn may be abbreviated $\nu = c\mu$.

Definition: Let μ and ν be two measures over the space (A,Σ); ν is uniform with respect to μ iff there is a positive real number c such that $\nu = c\mu$.

This is an equivalence relation among measures and implies that each is absolutely continuous with respect to the other. An equivalent way of stating the relation is that the density $d\nu/d\mu$ is equal to a positive real constant (μ-almost everywhere).

One recognizes, of course, that any such relation between disparate measures is at best an approximation. In general, it should not be taken literally at the microscopic level: A literally uniform distribution of land and water would just yield *mud* everywhere.

Induced Integrals

Let (A,Σ,μ), (B,Σ',μ') be two measure spaces. Let $f: A \to B$ be measurable, and such that, for all $E \in \Sigma'$,

$$\mu'(E) = \mu\{a \mid f(a) \in E\}. \tag{19}$$

(This says that μ' is induced by f from μ.) Finally, let g be a measurable, non-negative, extended real-valued function on B.

Theorem: Under the conditions stated,

$$\int_B g\, d\mu' = \int_A (g \circ f)\, d\mu. \tag{20}$$

Proof: A quick proof may be obtained by using the Young integral (16). In fact, from (19),

$$\mu'\{b \mid g(b) > t\} = \mu\{a \mid (g \circ f)(a) > t\}$$

so that

$$\int_B g\,d\mu' = \int_0^\infty \mu'\{b \mid g(b) > t\}dt = \int_0^\infty \mu\{a \mid (g\circ f)(a) > t\}dt$$

$$= \int_A (g\circ f)\,d\mu. \quad |||$$

Here $g\circ f$ is, of course, the composition of f and g. If g can take on negative values, then neither integral in (20) has yet been defined. However, to anticipate, it turns out that the theorem is still true in this case, in the sense that, if either integral in (20) is well-defined, then so is the other, and they are equal.

The elementary rules concerning "substitution of variables" in integration may be derived from (20).

There is a more general way of looking at relation (20). If we consider the indefinite integrals, $\int (g\circ f)d\mu$, and $\int g\,d\mu'$, then the latter is the measure induced on (B,Σ') by f from the former on (A,Σ).

Convergence Theorems

The following theorems are among the most useful in measure theory, and will be used repeatedly in this book. It is convenient to state them for integrands which are unrestricted in sign, even though we have so far only defined integration for non-negative integrands. (For the more general definition, see the section on signed measures below).

We distinguish formally between the sequence of extended real numbers, $(x_1, x_2, \ldots)$ — which is a family of numbers indexed by the integers 1, 2, ... — and the set $\{x_1, x_2, \ldots\}$ which is the range of this family. We have already defined the concepts of the supremum and infimum of a set of numbers. The sup and inf of the sequence $(x_1, x_2, \ldots)$ are defined simply as the sup and inf, respectively, of the set $\{x_1, x_2, \ldots\}$. Two slightly more complicated operations on sequences are needed here: lim sup and lim inf (limit superior and inferior).

Let $(x_1, x_2, \ldots)$ be a sequence of extended real numbers. Let $y_n = \sup \{x_n, x_{n+1}, \ldots\}$ for all $n = 1, 2, \ldots$. That is, y_n is the supremum of the numbers left in the sequence after deleting the first $n-1$ in order.

Definition: Lim sup of the sequence $(x_1, x_2, \ldots)$ is defined as the infimum of the set $\{y_1, y_2, \ldots\}$.

A similar construction reverses the roles of "inf" and "sup".

Definition: Let $z_n = \inf \{x_n, x_{n+1}, \ldots\}$ for all $n = 1, 2, \ldots$. Then $\sup \{z_1, z_2, \ldots\}$ is known as the lim inf of the sequence $(x_1, x_2, \ldots)$.

Definition: Sequence $(x_1, x_2, \ldots)$ converges to x_o iff the lim sup and lim inf of the sequence both equal x_o. x_o is the limit of the sequence and one writes: $x_n \to x_o$.

Examples.

(i) Let $(x_1, x_2, \ldots)$ be a sequence of real numbers. This converges to the real number x_o in the sense of this definition iff it converges to x_o in the ordinary sense of the term "converge".

(ii) The sequence (1, 0, 1, 0, 1, ...) has a lim sup of 1 and a lim inf of 0. (Proof: $y_n = 1$ and $z_n = 0$ for all n).

(iii) Let $(x_1, x_2, \ldots)$ be a non-decreasing sequence, and let $x_o = \sup \{x_1, x_2, \ldots\}$; then $(x_1, x_2, \ldots)$ converges to x_o. (Proof: $y_n = x_o$ for all n, hence $\inf \{y_1, y_2, \ldots\} = x_o$; $z_n = x_n$ for all n, hence $\sup \{z_1, z_2, \ldots\} = x_o$). From this example we note, e.g., that the sequence (1, 2, 3,...) converges to $+\infty$.

Now let $(f_1, f_2, \ldots)$ be a sequence of extended real-valued functions, with a common domain A.

Definition: Lim inf $(f_1, f_2, \ldots)$ is the function with domain A whose value at $a \in A$ equals lim inf $(f_1(a), f_2(a), \ldots)$. Lim sup $(f_1, f_2, \ldots)$ is defined analogously. If these two values are the same for all $a \in A$, the common function f thus determined is called the limit of the sequence $(f_1, f_2, \ldots)$, and one writes: $f_n \to f$.

One special case in which the limit exists is when the sequence is non-decreasing; that is, $f_n(a) \leq f_{n+1}(a)$ for all

$n = 1, 2, \ldots$, and all $a \in A$. This follows from example (iii) above, which also shows that for the limiting function f, $f(a)$ is the supremum of $\{f_n(a)\}$, $n = 1, 2, \ldots$, for all $a \in A$.

Theorem: (monotone convergence theorem) Let (A,Σ,μ) be a measure space; let (f_n), $n = 1, 2, \ldots$ be a non-decreasing sequence of measurable functions on A, with limit f. If

$$\int_A f_1 \, d\mu > -\infty, \tag{21}$$

then

$$\int_A f_n \, d\mu \to \int_A f \, d\mu. \tag{22}$$

Here we may note that the supremum of a sequence of measurable functions is measurable. Also condition (21) together with non-decreasingness, guarantees that all the integrals appearing in (22) are well-defined. It frequently happens that all the f_n's are non-negative, in which case (21) is automatically fulfilled.

Another version of the monotone convergence theorem uses infinite <u>series</u> rather than sequences.

Theorem: Let (A,Σ,μ) be a measure space; let (f_n), $n = 1, 2, \ldots$, be a sequence of non-negative measurable functions. Then

$$\int_A (f_1 + f_2 + \ldots) d\mu = \int_A f_1 \, d\mu + \int_A f_2 \, d\mu + \ldots \tag{23}$$

Here on the right we have an ordinary infinite series, whose sum is defined as usual as the limit of the partial sums. On the left the integrand is expressed as an infinite series of functions. This is to be understood pointwise: The value at point $a \in A$ is $f_1(a) + f_2(a) + \ldots$. Convergence is assured on both sides of (23) by non-negativity.

Equation (23) follows at once from the application of the monotone convergence theorem to the partial sums $f_1 + \ldots + f_n$, $n = 1, 2, \ldots$.

A closely related result involves an infinite series of measures. Let $\mu_1, \mu_2, \ldots$ all be measures on (A,Σ). The sum of the series

$$\mu_1 + \mu_2 + \ldots$$

is defined as the set function μ whose value at $E \in \Sigma$ is $\mu_1(E) + \mu_2(E) + \ldots$. One easily verifies that μ is a measure. We then have the following theorem.

Theorem: Let f be a non-negative measurable function on (A,Σ), and let $\mu_1 + \mu_2 + \ldots$ be a series of measures with sum μ. Then

$$\int_A f \, d\mu = \int_A f \, d\mu_1 + \int_A f \, d\mu_2 + \ldots .$$

Proof: Apply monotone convergence to the Young integral:

$$\int_A f\,d\mu = \int_0^\infty \mu\{a|f(a) > t\}dt = \int_0^\infty (\mu_1+\mu_2+\ldots)\{a|f(a) > t\}dt$$

$$= \int_0^\infty \mu_1\{a|f(a) > t\}dt + \ldots = \int_A f\,d\mu_1 + \ldots \quad |||$$

These last two theorems are used in the following section on product measures.

Theorem: (Fatou's lemma) Let (A,Σ,μ) be a measure space; let (f_n), $n = 1, 2, \ldots$ be a sequence of measurable functions on A such that $f_n \geq g$ for all n, g being another measurable function on A. If

$$\int_A g\,d\mu > -\infty, \tag{24}$$

then

$$\int_A (\liminf f_n)d\mu \leq \liminf \int_A f_n\,d\mu. \tag{25}$$

In (25), the "lim inf" on the left defines a function, which is to be integrated; the "lim inf" on the right, on the other hand, applies to the ordinary sequence of extended real numbers whose n-th term is $\int_A f_n\,d\mu$. We note that the lim inf of any sequence of measurable functions is measurable. Also condition (24) guarantees that all integrals appearing in (24) are well-defined. If all f_n's are non-negative, as is common, we may take $g = 0$, and (24) is automatically fulfilled.

An example will clarify the meaning of Fatou's lemma. Choose two sets $E_1, E_2 \in \Sigma$. Let f_n be the indicator function of E_1 if n is odd, and of E_2 if n is even. Lim inf f_n then equals $I_{E_1 \cap E_2}$; the sequence on the right of (25) is: $\mu(E_1)$, $\mu(E_2)$, $\mu(E_1)$,...; and, finally, (25) states that $\mu(E_1 \cap E_2) \leq$ minimum of $(\mu(E_1), \mu(E_2))$.

Theorem: (dominated convergence theorem). Let (A, Σ, μ) be a measure space; let (f_n), $n = 1, 2, \ldots$, be a sequence of measurable functions on A, with limit f; let $|f_n| \leq g$ for all n, g being another measurable function on A such that

$$\int_A g \, d\mu < \infty . \tag{26}$$

Then

$$\int_A f_n \, d\mu \to \int f \, d\mu. \tag{27}$$

The condition $|f_n| \leq g$, which states that the absolute value of f_n is dominated by g, may also be written: $-g(a) \leq f_n(a) \leq g(a)$ for all $a \in A$, and all $n = 1, 2, \ldots$. As above, the condition (26) guarantees that all integrals appearing in (27) are well-defined. (But even if all the integrals in (22), (25), or (27) are well-defined, one still cannot drop the conditions (21), (24), or (26), respectively, with impunity; this may be shown by counter-examples).

Extension of Set Functions

As we have already noted, a sigma-field is usually specified by mentioning a class of sets which generates it — e.g. the Borel field on the line is generated by the class of intervals. Similarly, a measure is often specified by stating its values just on some of the sets of its domain. For examples, Lebesgue measure is the one which assigns to each interval its ordinary length. As a second example, consider the product of (A,Σ,μ) and (B,Σ',ν): $\mu \times \nu$ is a measure whose value is specified in the first instance not on all of $\Sigma \times \Sigma'$ but only on the class of measurable rectangles.

Let $(\underline{A},\Sigma)$ be a measurable space, let $\mathcal{R}$ be a subclass of Σ, and let $\mu: \mathcal{R} \to$ extended reals be a set function defined on this subclass. The question arises: Does there exist a measure $\nu: \Sigma \to$ extended reals which coincides with μ on the latter's domain: $\nu(\underline{E}) = \mu(\underline{E})$ for all $\underline{E} \in \mathcal{R}$? In other words, can μ be <u>extended</u> to a measure on Σ? Furthermore, are there several such extensions or at most one?

We now specify certain conditions on μ and $\mathcal{R}$ which enable us to answer such questions.

<u>Definition</u>: Set function $\mu: \mathcal{R} \to$ non-negative extended reals is <u>countably</u> <u>additive</u> iff, for any countable packing $\underline{E}_1, \underline{E}_2, \ldots$ of $\mathcal{R}$-sets whose union $\underline{E}$ is also an $\mathcal{R}$-set, we have

$$\mu(\underline{E}) = \mu(\underline{E}_1) + \mu(\underline{E}_2) + \ldots \, .$$

This is a slight generalization of the concept of countable additivity on a sigma-field: The condition that $E \in \mathcal{R}$ must be stated explicitly, since $\mathcal{R}$ is not necessarily closed under countable unions.

Definition: Collection $\mathcal{R}$ is a semi-ring iff

(i) $\emptyset \in \mathcal{R}$; and

(ii) if $E, F \in \mathcal{R}$, then $E \cap F \in \mathcal{R}$, and

(iii) if $E, F \in \mathcal{R}$, and $E \subseteq F$, then there is a finite sequence $G_0, \ldots, G_n$ of $\mathcal{R}$-sets such that $G_0 = E$, $G_n = F$, $G_{i-1} \subseteq G_i$, and $G_i \setminus G_{i-1} \in \mathcal{R}$, $i = 1, \ldots, n$.

For example, the collection of all intervals on the real line (together with $\emptyset$) is a semi-ring. The collection of measurable rectangles in a product space (and $\emptyset$) is a semi-ring.

Theorem: Let (A, Σ) be a measurable space, let $\mathcal{R}$ be a semi-ring which generates Σ, and let non-negative $\mu: \mathcal{R} \to$ extended reals be countably additive, with $\mu(\emptyset) = 0$. Then there exists a measure ν with domain Σ which extends μ.

If, in addition, there is a countable collection $\mathcal{G} \subseteq \mathcal{R}$ which covers A such that $\mu(G) < \infty$ all $G \in \mathcal{G}$, then there is exactly one such extension.

The premise that $\mathcal{R}$ is a semi-ring can be weakened considerably without invalidating these conclusions.

Definition: Collection $\mathcal{R}$ is a <u>weak semi-sigma-ring</u> iff (i) $\emptyset \in \mathcal{R}$; and (ii) for all $E, F \in \mathcal{R}$, $E \setminus F$ can be partitioned into a countable number of $\mathcal{R}$-sets. $\mathcal{R}$ is a <u>semi-sigma-ring</u> iff it is a weak semi-sigma-ring, and for all $E, F \in \mathcal{R}$, $E \cap F$ can also be partitioned into a countable number of $\mathcal{R}$-sets.

Any semi-ring is a semi-sigma-ring. For let $E, F \in \mathcal{R}$; then $E \cap F \in \mathcal{R}$, so the collection consisting of $E \cap F$ alone is a countable partition of $E \cap F$ into $\mathcal{R}$-sets; furthermore, the collection $\{G_1 \setminus G_0, G_2 \setminus G_1, \ldots, G_n \setminus G_{n-1}\}$, where $G_0 = E \cap F$, $G_n = E$, $G_{i-1} \subseteq G_i$, $i = 1, \ldots, n$, is a finite, hence countable, partition of $E \setminus F$ into $\mathcal{R}$-sets.

This shows that the following theorem is a generalization of the preceding one.

Theorem: Let (A, Σ) be a measurable space, let $\mathcal{R}$ be a <u>semi-sigma-ring</u> which generates Σ, and let non-negative $\mu: \mathcal{R} \to$ extended reals be countably additive with $\mu(\emptyset) = 0$. Then there exists a measure ν with domain Σ which extends μ.

Instead, let $\mathcal{R}$ be a <u>weak semi-sigma-ring</u> which generates Σ; let ν_1, ν_2 be two measures on Σ which coincide on $\mathcal{R}$-sets; and let there be a countable collection $\mathcal{G} \subseteq A$ which covers A such that $\nu_1(G) = \nu_2(G) < \infty$, all $G \in \mathcal{G}$. Then $\nu_1 = \nu_2$ throughout Σ.

The proof of this theorem is very long and would take us too far afield to set down here.

Abcont and Product Measures

Let $(\underline{A},\Sigma,\mu)$ and $(\underline{B},\Sigma',\nu)$ be two measure spaces. We have already defined the concept of the product, $(\underline{A} \times \underline{B}, \Sigma \times \Sigma')$, of the two measurable spaces, $(\underline{A},\Sigma)$ and $(\underline{B},\Sigma')$. We now define the notion of a measure, λ, on the product space which is in a sense the "product" of the measures μ and ν.

Definition: Measure λ on $(\underline{A} \times \underline{B}, \Sigma \times \Sigma')$ is a generalized product of μ and ν iff

$$\lambda(\underline{E} \times \underline{F}) = \mu(\underline{E}) \cdot \nu(\underline{F}) \qquad (28)$$

for all $\underline{E} \in \Sigma$ and $\underline{F} \in \Sigma'$.

Intuitively, (28) says that the λ-measure of any "rectangle" $\underline{E} \times \underline{F}$ is the ordinary product of the measures of its "sides". (In evaluating the right-hand side, remember that $0\cdot\infty = 0$). As an example, let μ and ν both be Lebesgue measure (= length) on the real line, and let λ be (two-dimensional) Lebesgue measure (= area) on the plane. Then (28) is satisfied: The area of a rectangle is the product of its sides.

Does a generalized product exist for any pair of measures? Yes it does. (This may be proved via the monotone convergence theorem coupled with the extension theory just discussed.) Is it unique? No, not always; But if μ and ν are both sigma-finite, then uniqueness is guaranteed. (Thus, two-dimensional Lebesgue measure is the only possible product of Lebesgue measure with itself).

We are interested ~~in this section~~ only in ~~those~~ product measures ~~which~~ can be expressed as integrals. Choose $\underline{G} \in (\Sigma \times \Sigma')$, and consider the following expression:

$$\lambda(\underline{G}) = \int_{\underline{A}} \nu\{\underline{b} \mid (\underline{a},\underline{b}) \in \underline{G}\}\mu(d\underline{a}). \qquad (29)$$

For a given point $\underline{a}_o \in \underline{A}$, the set $\{\underline{b} \mid (\underline{a}_o,\underline{b}) \in \underline{G}\}$ is the right $\underline{a}_o$-section of $\underline{G}$. We know this is a measurable subset of $\underline{B}$, hence it has a ν-value. We associate this value with point $\underline{a}_o$. We now have a well-defined function with domain $\underline{A}$, taking values in the non-negative extended real ~~numbers~~; this function is precisely the integrand in (29). The integral with respect to μ may now be taken, provided the integrand just defined is measurable with respect to $(\underline{A},\Sigma)$. Suppose this to be the case for each $G \in (\Sigma \times \Sigma')$. Then (29) defines a set function λ.

λ is ~~easily shown to be~~ a measure (use monotone convergence in its additive form). Furthermore, letting $\underline{G} = \underline{E} \times \underline{F}$, we obtain

$$\lambda(\underline{E} \times \underline{E}) = \int_{\underline{E}} \nu(\underline{F})\, d\mu = \mu(\underline{E}) \cdot \nu(\underline{F}),$$

so that λ is a generalized product of μ and ν.

Thus the expression (29) yields a generalized product, provided the sticky issue of measurability is resolved. If ν is finite, it is known that the integrand in (29) is measurable for all $\underline{G} \in (\Sigma \times \Sigma')$. To generalize this result, we now introduce a new class of measures, the abcont measures.

Theorem: Let (A,Σ,μ) be a measure space. If μ has any of the following four properties, then it has all of them:

(1) there is a series of finite measures ν_n, $n = 1, 2, \ldots$, on (A,Σ) such that

$$\mu = \nu_1 + \nu_2 + \ldots \quad (30)$$

(2) there is a measure space (B,Σ',ν), ν sigma-finite, and a measurable function $f: B \to A$, such that μ is the measure induced by f from ν;

(3) there is a finite measure ν on (A,Σ) with respect to which μ is absolutely continuous ($\mu << \nu$);

(4) there is a finite measure ν on (A,Σ), and a measurable function, $f: A \to$ non-negative extended real numbers, such that

$$\mu = \int f \, d\nu.$$

Proof: (1) implies (2): Let $N = \{1, 2, \ldots\}$, let Σ'' be the class of all subsets of N, let

$$(B,\Sigma') = (A \times N, \Sigma \times \Sigma''),$$

let $f: B \to A$ be the projection, $f(a,n) = a$, and define ν as follows: for any set of the form $E \times \{n\}$, where $E \in \Sigma$, $n = 1, 2, \ldots$, let $\nu(E \times \{n\}) = \nu_n(E)$. Any set $G \in \Sigma'$ is a countable disjoint union of sets of this form, so measure ν is fully determined on Σ'. ν is sigma-finite, since $\nu(A \times \{n\}) = \nu_n(A) < \infty$, $n = 1, 2, \ldots$, and the sets $A \times \{n\}$ partition B.

Also, for any $\underline{E} \in \Sigma$,

$$\nu\{\underline{b}|\underline{f}(\underline{b}) \in \underline{E}\} = \nu(\underline{E} \times \underline{N}) = \nu(\underline{E} \times \{1\}) + \nu(\underline{E} \times \{2\}) + \ldots$$

$$= \nu_1(\underline{E}) + \nu_2(\underline{E}) + \ldots = \mu(\underline{E}).$$

Thus μ is indeed the measure induced by $\underline{f}$ from ν.

(2) implies (3): If $\mu = 0$ this is trivial; hence we may assume $\mu \neq 0$, so that $\nu \neq 0$. Let $\{\underline{B}_1, \underline{B}_2, \ldots\}$ be a countable measurable partition of $\underline{B}$ such that $0 < \nu(\underline{B}_n) < \infty$, for all $\underline{n}$, and define the set function λ on $(\underline{A}, \Sigma)$ as follows. For $\underline{E} \in \Sigma$, $\lambda(\underline{E})$ is the summation of

$$2^{-\underline{n}} \nu\left[\underline{B}_n \cap \{\underline{b}|\underline{f}(\underline{b}) \in \underline{E}\}\right] / \nu(\underline{B}_n) \tag{31}$$

over $\underline{n} = 1, 2, \ldots$. Each of these terms defines a measure on $(\underline{A}, \Sigma)$, hence λ itself is such a measure. Furthermore, $\lambda(\underline{A}) = 2^{-1} + 2^{-2} + \ldots = 1$, so λ is finite. It remains to show that $\mu << \lambda$. Suppose $\lambda(\underline{E}) = 0$; then each of the terms (31) vanishes so that

$$\nu\left[\underline{B}_n \cap \{\underline{b}|\underline{f}(\underline{b}) \in \underline{E}\}\right] = 0,$$

all $\underline{n} = 1, 2, \ldots$. Adding over $\underline{n}$, we obtain $\nu\{\underline{b}|\underline{f}(\underline{b}) \in \underline{E}\} = 0$; since μ is induced by $\underline{f}$ from ν, it follows that $\mu(\underline{E}) = 0$; hence $\mu << \lambda$.

(3) implies (4): Apply the Radon-Nikodym theorem.

(iv) implies (i): For each $n = 1, 2, \ldots$ define $f_n : A \to$ reals by

$$f_n(a) = 0 \qquad \text{if } f(a) \leq n - 1;$$

$$f_n(a) = f(a) - (n - 1) \qquad \text{if } n - 1 < f(a) < n;$$

$$f_n(a) = 1 \qquad \text{if } f(a) \geq n.$$

Each f_n is non-negative and measurable. Define ν_n by

$$\nu_n = \int f_n \, d\nu.$$

Since ν and f_n are bounded, each ν_n is bounded. Finally,

$$\mu = \int f \, d\nu = \int [f_1 + f_2 + \ldots] \, d\nu$$

$$= \int f_1 \, d\nu + \int f_2 \, d\nu + \ldots = \nu_1 + \nu_2 + \ldots,$$

establishing (30). (The third equality follows from monotone convergence.)

We now have a closed circle of implications, hence these four properties are logically equivalent. |||

Definition: Measure μ is <u>abcont</u> iff it satisfies any (hence all) of the foregoing properties.

"Abcont" is an acronym for "absolutely continuous" and is suggested by property (iii). But keep in mind that absolute continuity is a relation between <u>two</u> measures, while abcontness is a property of a <u>single</u> measure.

Any sigma-finite measure is abcont. This is clear from property (ii), for if μ is sigma-finite we may take $(\underline{B},\Sigma',\nu) = (\underline{A},\Sigma,\mu)$ and let $\underline{f}$ be the identity mapping. Also, in property (iv), if we impose the additional condition that $\underline{f}$ be <u>finite</u>, we have a characterization of sigma-finite measures. This also yields a decomposition for abcont measures: When restricted to the set $\{\underline{a}|\underline{f}(\underline{a}) < \infty\}$, μ is sigma-finite, and when restricted to the set $\{\underline{a}|\underline{f}(\underline{a}) = \infty\}$, μ takes on just two possible values, 0 and ∞.

There exist abcont measures which are not sigma-finite. In fact, to produce such a measure simply take any finite measure $\nu \neq 0$, and set $\mu(\underline{E}) = \infty$ whenever $\nu(\underline{E}) > 0$, and $\mu(\underline{E}) = 0$ whenever $\nu(\underline{E}) = 0$. μ is easily seen to be a measure; that it is abcont follows from the observation that

$$\mu = \nu + \nu + \nu + \ldots$$

(property (i)), or that $\mu << \nu$ (property (iii)), or that $\mu = \int(\infty)d\nu$ (property (iv)).

Abcont measures are of much less importance to us than the narrower class of sigma-finite measures. There are two reasons for introducing them. First, even if one is interested only in sigma-finite measures, abcont non-sigma-finite measures may appear as the result of perfectly straightforward operations, e.g., induction as in property (7). For example, one or both marginals of a sigma-finite measure ν on a product space $A \times B$ may have this property. (Exercise: Verify that the marginals of two-dimensional Lebesgue measure on the plane are abcont non-sigma-finite).

Second, in many cases abcontness is a more natural assumption than sigma-finiteness, in that it yields results that are both stronger and more transparent, with proofs that are clearer. This is especially the case in the present section on product measures, and happens from time to time throughout the book.

Each of the four properties of abcont measures yields a "closure" theorem. We gather these results under one roof:

Theorem: (1) The sum of a series of abcont measures on (A,Σ) is abcont;

(2) let (B,Σ',ν) and (A,Σ,μ) be measure spaces, and let measurable $f: B \to A$ induce μ from ν; if ν is abcont, μ is abcont;

(3) if $\mu \ll \nu$, and ν is abcont, then μ is abcont;

(4) let measure ν on (A,Σ) be abcont, and let $f: A \to$ extended reals be non-negative measurable; then $\mu = \int f\, d\nu$ is abcont.

Proof: (1) Let ν_m be abcont, $m = 1, 2, \ldots$, so that

$$\nu_m = \lambda_{m1} + \lambda_{m2} + \ldots$$

for some finite measures λ_{mn}, $m,n = 1, 2, \ldots$. The sum of the ν_m's is then the sum of the double series λ_{mn}. The λ's being countable and non-negative, they may be rearranged in a single series. Hence summation ν_m is abcont by property (1).

(2) Since ν is abcont, there exists a measure space $(\underline{C},\Sigma'',\lambda)$, λ sigma-finite, and a measurable function $g: \underline{C} \to \underline{B}$ inducing ν from λ. But then $(f \circ g): \underline{C} \to \underline{A}$ induces μ from λ, so that μ is abcont by property (2).

(3) Since ν is abcont, there is a finite measure λ for which $\nu << \lambda$. But then $\mu << \lambda$, so μ is abcont by property (3).

(4) Since ν is abcont, there is a finite measure λ and a non-negative measurable $g: \underline{A} \to$ extended reals such that $\nu = \int g\, d\lambda$. But then

$$\mu = \int f\, d\left[\int g\, d\lambda\right] = \int (fg)\, d\lambda,$$

so μ is abcont by property (4).

Let us now return to the problem of constructing product measures. Consider again the integral expression (29), and suppose now that ν is abcont. The first property of abcont measures -- that they are the sums of series of finite measures -- is the key to the discussion from here on. Thus we may write $\nu = \nu_1 + \nu_2 + \ldots$, where each ν_n is finite. Now, from what was said above, the function on $\underline{A}$ given by

$$f_n(a) = \nu_n\{b \mid (a,b) \in G\}$$

is measurable for each $n = 1, 2, \ldots$; hence the sum of all of them is a measurable function. That is, the abcontness assumption on ν guarantees that the integrand in (29) is measurable, so that the integral expression (29) yields a well-defined product measure. Note that no restriction on μ need be imposed.

Next, suppose ~~that~~ μ and ν are both abcont. In this case it is possible to reverse their roles, ~~which~~ yields the expression

$$\tilde{\lambda}(G) = \int_B \mu\{a \mid (a,b) \in G\}\, \nu(db). \quad (32)$$

We now have two product measures, λ and $\tilde{\lambda}$. ~~It turns out,~~ however, ~~that~~ these are identical. To show this, we start from the observation made above that two finite (in fact, ~~sigma~~-finite) measures have a unique product measure. Let ν be decomposed as above, and, similarly, write

$$\mu = \mu_1 + \mu_2 + \ldots,$$

where all measures μ_m are finite. Substituting μ_m and ν_n for μ and ν, respectively, in (29) and (32), we obtain two measures ~~which~~ may be written λ_{mn} and $\tilde{\lambda}_{nm}$. We must have

$$\lambda_{mn} = \tilde{\lambda}_{nm}$$

since both of these are products of the finite measures μ_m and ν_n. Taking the double summation over $m,n = 1, 2, \ldots,$ and applying two versions of monotone convergence, we arrive finally at

$$\lambda = \tilde{\lambda}.$$

Definition: Let (A,Σ,μ) and (B,Σ',ν) be two measure spaces, at least one of which is abcont. $(\mu \times \nu)$ is the product measure obtained from (29) (if μ is abcont) or from (32) (if ν is abcont).

If both are abcont then of course either integral formula may be used, yielding the same result. From here on we shall refer to $\mu \times \nu$ simply as the product of μ and ν. (It should be noted, however, that there may exist generalized products of μ and ν other than $\mu \times \nu$. Indeed, one such case occurs with μ abcont and ν bounded, but we shall not pause to examine this counter-example).

If μ and ν are both finite, then $\mu \times \nu$ is finite (since $\mu(A) \cdot \nu(B) < \infty$). If μ and ν are both sigma-finite, then $\mu \times \nu$ is sigma-finite (countably partition A and B so that μ and ν are finite on each respective piece; $\mu \times \nu$ is then finite on each $A_m \times B_n$). Finally, if μ and ν are both abcont, then $\mu \times \nu$ is abcont (to see this, note that each λ_{mn} above is finite). (However, in this case there sometimes exist other generalized products of μ and ν which are not abcont).

Since $\mu \times \nu$ is a measure, we may integrate with respect to it. The following result is important.

Theorem: Let (A,Σ,μ) and (B,Σ',ν) be measure spaces, with ν abcont; let $f: A \times B \to$ extended reals be measurable and non-negative. Then

$$\int_{A\times B} f\, d(\mu \times \nu) = \int_A \mu(da) \int_B \nu(db)\, f(a,b)\,. \tag{33}$$

The right-hand side of (33) is to be interpreted as follows. For fixed $a_o \in A$, $f(a_o,\cdot)$ has domain B. Integrating with respect to ν, we obtain a number, which depends on the point a_o. The resulting function with domain A is measurable, and may be integrated with respect to μ. Thus (33) expresses a single integral with respect to product measure in terms of an iterated integral with respect to the component measures. Results of this kind go under the name of Fubini's theorem (or Fubini-Tonelli's theorem).

Two observations are worth making. First, let $G \in \Sigma \times \Sigma'$, and let $f = I_G$, the indicator function of set G. Then one easily verifies that (33) reduces to (29), the defining equation for product measure. Second, suppose μ and ν are both abcont. Then the rôles of μ and ν may be reversed. Combining this version of Fubini's theorem with the one above, we see that the same result is obtained independent of the order in which an iterated integration is carried out (provided the integrand is non-negative).

The proof of Fubini's theorem may be outlined. For f an indicator function the result is immediate, as noted above. A simple function is a weighted sum of indicators, and the result then follows for f simple. Finally, noting that any non-negative measurable function is the pointwise limit of a non-decreasing sequence of simple functions, we apply monotone convergence to the result just obtained, yielding the general theorem.

We ~~want to~~ generalize these results in two directions: first, to the product of more than two spaces; second, from ordinary to conditional measures. The second direction is of greater importance for us, and we start with it.

Definition: Let (A,Σ) and (B,Σ') be measurable spaces; a conditional measure is a function $\nu: A \times \Sigma' \rightarrow$ extended reals such that

(i) for each $a_o \in A$, the right section $\nu(a_o,\cdot)$ is a measure on (B,Σ'); ~~and~~

(ii) for each $E \in \Sigma'$ the left section $\nu(\cdot,E)$ is a measurable function on (A,Σ).[34]

Note the peculiar domain of ν: the cartesian product of the universe set A of one space and the ~~sigma~~ σ-field Σ' of the other space. Thus ν assigns numbers to pairs (a,E) consisting of a point and a set.

Suppose we have a conditional measure ν as above, ~~together with~~ and an ordinary measure μ on (A,Σ). Let $G \in \Sigma \times \Sigma'$ (this is the product sigma-field, not the cartesian product), and consider the expression

$$\lambda(G) = \int_A \nu(a, \{b \mid (a,b) \in G\}) \, \mu(da). \qquad (34)$$

This is the same as (29) except for the extra "a" in the argument of ν. This causes no complications. As before, for given G, the expression $\nu(a, \{b \mid (a,b) \in G\})$ defines a function with domain A, which is to be integrated with respect to μ. If the integrand is measurable for each $G \in \Sigma \times \Sigma'$, then λ is a well-defined set function; in fact, λ is a measure, as one verifies by applying monotone convergence.

A known sufficient condition for the integrand in (34) to be measurable is that ν be a finite conditional measure. To generalize, we have to extend the abcontness concept.

Definition: Conditional measure $\nu: A \times \Sigma' \rightarrow$ extended reals is abcont iff there exists a series of finite conditional measures ν_n, n = 1, 2, ..., such that

$$\nu = \nu_1 + \nu_2 + \ldots \qquad (35)$$

Here (35) is to be understood as usual in the pointwise sense: For the pair $(a,E) \in A \times \Sigma'$, the values $\nu_n(a,E)$ constitute a numerical series whose sum is $\nu(a,E)$ (both sides may equal $+\infty$).

We now proceed as above. For given $G \in \Sigma \times \Sigma'$, the expression

$$\nu_n(a, \{b \mid (a,b) \in G\})$$

defines a measurable function on A for each n; hence the sum of all these functions — which is the integrand in (34) — is measurable. Thus $\lambda(G)$ is a well-defined measure.

We refer to this as the product of μ and ν, and denote it as above by $\mu \times \nu$. Here μ may be any measure on (A,Σ), while ν is an abcont conditional measure with domain $A \times \Sigma'$.

This generalizes the preceding construction, if we identify the ordinary measure ν on (B,Σ') with the conditional measure $\tilde{\nu}$ given by

$$\tilde{\nu}(a,G) = \nu(G). \tag{36}$$

In other words, a conditional measure which is independent of its first argument may be thought of as an ordinary measure with domain Σ', in which case the formula (34) reduces to (29). Note however that by taking ν conditional we have lost the summetry between μ and ν. No reversal of roles is possible, even if μ is abcont. Note also that with the identification (36), conditional abcontness reduces to ordinary abcontness.

Fubini's theorem remains valid if ν is taken to be abcont conditional: Simply insert the extra argument "a" in (33).

We now generalize to more than two spaces. Let measurable spaces $(A_1, \Sigma_1), \ldots, (A_n,\Sigma_n)$ be given. We are also given functions $\mu_1, \ldots, \mu_n$; μ_i has domain $(A_1 \times \ldots \times A_{i-1}) \times \Sigma_i$, and is a conditional measure. That is, for a given point

$(a_1, \ldots, a_{i-1})$ in the product space $A_1 \times \ldots \times A_{i-1}$, the right section $\mu_i(a_1, \ldots, a_{i-1}, \cdot)$ is a measure on (A_i, Σ_i); and, for a given set $E \in \Sigma_i$, the left section $\mu(\cdot, \cdot, \ldots, \cdot, E)$ is a measurable function on $(A_1 \times \ldots \times A_{i-1}, \Sigma_1 \times \ldots \times \Sigma_{i-1})$. ($\mu_1$ is just an ordinary measure with domain Σ_1). Finally, let a non-negative measurable function $f: A_1 \times \ldots \times A_n \to$ extended reals be given, and consider the expression

$$\int_{A_1} \mu_1(da_1) \int_{A_2} \mu_2(a_1, da_2) \int_{A_3} \ldots \int_{A_n} \mu_n(a_1, \ldots, a_{n-1}, da_n) f(a_1, \ldots, a_n). \tag{37}$$

These iterated integrals are to be evaluated from right to left. That is, first fix $a_1, \ldots, a_{n-1}$, and integrate the right-section of f with respect to the right section of μ_n over A_n. This yields a function with domain $A_1 \times \ldots \times A_{n-1}$. Next, fix $a_1, \ldots, a_{n-2}$ and integrate this function with respect to μ_{n-1} over A_{n-1}, to obtain a function with domain $A_1 \times \ldots \times A_{n-2}$. Continue, finishing with an integration over A_1 with respect to μ_1.

For this process to be well-defined we must obtain a measurable integrand at each stage. What conditions guarantee this? We start with the case where f is an indicator function. Let $G \in (\Sigma_1 \times \ldots \times \Sigma_n)$ (This is the product sigma-field, not the cartesian product), and let $f = I_G$.

Theorem: If $\mu_2, \mu_3, \ldots, \mu_n$ are all abcont conditional measures, then the iterated integral is well-defined for $f = I_G$, for any $G \in (\Sigma_1 \times\ldots\times \Sigma_n)$. The resulting set function $\lambda(G)$ is, in fact, a measure on $(A_1 \times\ldots\times A_n, \Sigma_1 \times\ldots\times \Sigma_n)$.

Definition: $\lambda(G)$ is called the product measure of $\mu_1, \ldots, \mu_n$, and written $\mu_1 \times\ldots\times \mu_n$.

Note that no restriction need be placed on μ_1. Note also that the conditioning structure permits no rôle reversals: The successive integrations must be performed in the specified order. A special case arises when the µ's are actually independent of their point-arguments; if so, each μ_i may be thought of as an ordinary measure on (A_i, Σ_i). If we let G be the measurable rectangle $E_1 \times\ldots\times E_n$, we easily obtain in this case

$$(\mu_1 \times\ldots\times \mu_n)(E_1 \times\ldots\times E_n) = \mu_1(E_1)\cdot\mu_2(E_2)\cdots\mu_n(E_n),$$

generalizing the product measure relation to n components. For example, n-dimensional Lebesgue measure is the product of n one-dimensional Lebesgue measures.

We next generalize Fubini's theorem:

Theorem: Let $\mu_2, \ldots, \mu_n$, and f be as above. Then (37) is well-defined, and equals

$$\int_{A_1 \times \ldots \times A_n} f \, d(\mu_1 \times \ldots \times \mu_n).$$

This and the preceding theorem are best proved simultaneously by induction on n, the number of components. The case $n = 2$ has already been discussed, and these results may be used to go from n to $n + 1$. We omit details.

Finally, we make some remarks about the properties of product measures. If μ_1 is abcont (as well as $\mu_2, \ldots, \mu_n$ being conditionally abcont), then $\mu_1 \times \ldots \times \mu_n$ is abcont. This may be proved by induction on n. If $\mu_1, \mu_2, \ldots, \mu_n$ are all <u>bounded</u>, then $\mu_1 \times \ldots \times \mu_n$ is bounded. (Note that boundedness is a stronger condition than finiteness for conditional measures; for ordinary measures the two concepts of course coincide). What about the intermediate case of sigma-finiteness? For this we need one more concept.

Definition: Given (A,Σ) and (B,Σ'), let $\nu: A \times \Sigma' \to$ extended reals be a conditional measure; ν is <u>uniformly sigma-finite</u> iff there is a countable collection $G \subseteq \Sigma'$ such that, for all $(a,b) \in A \times B$, there is a set $E \in G$ such that $b \in E$ and $\nu(a,E) < \infty$.

This property is a little stronger than mere sigma-finiteness of each right section $\nu(a,\cdot)$. It implies conditional abcontness. As an example, let ν be independent of its point-argument; then ν is uniformly sigma-finite iff it is sigma-finite when thought of as an ordinary measure.

Theorem: Let μ_1 be sigma-finite, and let $\mu_2, \ldots, \mu_n$ be uniformly sigma-finite; then $\mu_1 \times \ldots \times \mu_n$ is sigma-finite.

Proof: First take the case $n = 2$: μ_2 is uniformly sigma-finite, so there are Σ_2-sets $F_1, F_2, \ldots$ such that, for any (a_1, a_2), there is an F_i for which $a_2 \in F_i$ and $\mu_2(a_1, F_i) < \infty$. Also there is a covering $E_1, E_2, \ldots$ of A_1 such that $\mu_1(E_j) < \infty$, all j, by μ_1 sigma-finite. Define

$$G_{ijk} = \{a_1 | a_1 \in E_j \text{ and } \mu_2(a_1, F_i) < k\} \times F_i,$$

for all $i, j, k = 1,2,3,4, \ldots$. These sets G_{ijk} form a countable measurable covering of $A_1 \times A_2$. Measurability follows from the fact that $\mu_2(\cdot, F_i)$ is measurable; also, any $(a_1, a_2) \in (E_j \times F_i)$ with $\mu_2(a_1, F_i) < \infty$ for some i, j, and so $(a_1, a_2) \in G_{ijk}$ for some $k = 1, 2, \ldots$, proving the covering property. Finally,

$$(\mu_1 \times \mu_2)(G_{ijk}) \leq k \cdot \mu_1(E_j) < \infty,$$

since the integrand has k for an upper bound, and is zero outside E_j. This proves $\mu_1 \times \mu_2$ is sigma-finite.

For the general case, proceed by induction on n: Assume true for $n - 1$, so that $\mu_1 \times \ldots \times \mu_{n-1}$ is sigma-finite. From Fubini's theorem we obtain

$$\mu_1 \times \ldots \times \mu_n = (\mu_1 \times \ldots \times \mu_{n-1}) \times \mu_n.$$

But this expresses the n-fold product as a 2-fold product, of which the left component is sigma-finite by induction hypothesis, and the right component, μ_n, is uniformly sigma-finite. Hence $\mu_1 \times \ldots \times \mu_n$ is sigma-finite, completing the induction. ||| □■

Distribution Functions[35]

Our entire discussion of measure theory has been framed so as to apply to measurable spaces in general. The present topic will be the one exception, in that the concepts apply only to finite products of the real line — that is, to n-space, the set of all n-tuples of real numbers, with the corresponding n-dimensional Borel field. This measurable space will be denoted (A^n, Σ^n) in the present discussion.

Definition: Let μ be a measure on (A^n, Σ^n), and let f be a real-valued function with domain A^n. f is a distribution function for μ (in the narrow sense) iff, for every n-tuple of real numbers, $(b_1, \ldots, b_n)$, we have

$$f(b_1, \ldots, b_n) = \mu\{(x_1, \ldots, x_n) \mid x_i < b_i, \text{ for all } i = 1, \ldots, n\} \quad (38)$$

For $n = 1$, the indicated set on the right is a half-way $\{x \mid -\infty < x < b_1\}$. For $n = 2$, it is a "southwest" quadrant of the plane. There is also a "wide-sense" definition:

Definition: Let μ be a measure on (A^n, Σ^n) and f a real-valued function on A^n. f is a distribution function for μ (in the wide sense) iff for every n pairs of real numbers (a_i, b_i), $i = 1, \ldots, n$, with $a_i < b_i$, the value

$$\mu\{(x_1, \ldots, x_n) \mid a_i \le x_i < b_i,\ i = 1, \ldots, n\} \qquad (39)$$

is equal to the following sum of 2^n terms:

$$f(b_1, \ldots, b_n) - f(a_1, b_2, \ldots, b_n) - f(b_1, a_2, b_3, \ldots, b_n) - \ldots + f(a_1, a_2, b_2, \ldots, b_n) + \ldots + (-1)^n f(a_1, \ldots, a_n). \qquad (40)$$

(In (40) we run over terms of the form $f(y_1, \ldots, y_n)$ where $y_i = a_i$ or b_i, and all 2^n possible selection patterns are used; if an even number of a_i's appear, the term enters with a "+" sign; if an odd number, with a "-" sign).

For $n = 1$, the set appearing in (39) is an interval, including its left but not its right endpoint. For $n = 2$, the set is an ordinary rectangle, including two of its four sides and one of its four corners. It is convenient to refer to sets in general of the form (39) as bounded intervals. Every bounded interval in n-space has 2^n corners, and f is a distribution function for μ iff the measure of any interval is equal to the sums and differences of the values of f on these corners according to the sign rule stated above.

~~Let us give~~ some examples follow.

(i) $\mu = 0$ and $f = 0$ identically; then f is a distribution function for μ in both senses.

(ii) $\mu = 0$, and f is constant $\neq 0$; then f is a distribution function in the wide but not the narrow sense. (Proof: there are an equal number of "+" and "-" terms in (40), so the sum is zero.)

(iii) $\mu(E) = 1$ if $(0,\ldots,0) \in E$, and $= 0$ otherwise; the function f for which $f(b_1,\ldots,b_n) = 1$ if all b_i's are positive, and = 0 otherwise, is a distribution function in both senses.

(iv) Let μ be n-dimensional Lebesgue measure; this has no distribution function in the narrow sense, but the function $f(b_1,\ldots,b_n) = b_1 b_2 \ldots b_n$ is one in the wide sense. (Proof: Lebesgue measure in (39) is the product $(b_1-a_1)(b_2-a_2)\ldots(b_n-a_n)$; when multiplied out we get 2^n terms which are exactly the terms of ~~(75)~~ (40) with the proper signs.)

We want answers to the following questions. Given μ, when does it have a distribution function in either sense, and are these unique? Given f, when is it the distribution function, in either sense, of some measure, and is this measure unique?

A partial answer can be given immediately. ~~It is obvious~~ from (38) it can be seen that a measure has at most one distribution function in the narrow sense, and that it does have such a function iff it is finite on every set of the type appearing in (38).

The corresponding result for wide-sense distribution functions is a little more difficult.

Theorem: Measure μ has a distribution function in the wide sense iff it is finite on every bounded interval. If f is a wide-sense distribution function for μ, then g is another one iff $g - f$ is of the form $h_1 + \ldots + h_n$, where $h_i : A^n \to$ reals does not depend on its i-th coordinate.

(That is, a change in the i-th coordinate produces no change in $h_i(x_1, \ldots, x_n)$, whatever the values of the other $n - 1$ coordinates).

Thus if f is a wide-sense distribution function for μ, one can add any real-valued function of $n - 1$ variables to f_1 and still retain that property. For $n = 1$, a constant may be added. For $n = 2$, $g(x_1, x_2) = f(x_1, x_2) + b_1(x_1) + b_2(x_2)$ is a wide-sense distribution function for μ if f is. The reason is that any such additions cancel out in the differencing process (40).

If μ is finite on every bounded interval, one can write an explicit formula for f, namely

$$f(b_1, \ldots, b_n) = \pm \mu\{(x_1, \ldots, x_n) | 0 \leq x_i < b_i \text{ or } b_i \leq x_i < 0, \quad i = 1, \ldots, n\}. \tag{41}$$

(Here the condition $0 \leq x_i < b_i$ is to be imposed if b_i is positive, the condition $b_i \leq x_i < 0$ if b_i is negative; the "+" is to be taken if the number of negative b_i's is even, the "-" sign if that number is odd; finally, $f(b_1, \ldots, b_n) = 0$ if any of the b_i's equals zero). As an example, for Lebesgue measure (41) comes out to $f(b_1,\ldots,b_n) = b_1b_2\cdots b_n$, a distribution function already referred to. The general wide-sense distribution function is then obtained by adding arbitrary functions $h_1+\ldots+h_n$ to (41), as in the theorem above.

The various conditions imposed on μ have the following relations. If μ is finite on every set of the form (38), then μ is finite on every bounded interval; and this in turn implies that μ is sigma-finite. However, as one may show by examples, neither of these two implications can be reversed.

Next we come to the relation between wide- and narrow-sense:

Theorem: f is a distribution function for μ in the narrow sense iff it is a distribution function for μ in the wide sense and, for any $i_o = 1,\ldots,n$, and for any $n - 1$ fixed real numbers b_i, ($i = 1,\ldots, n$; $i \neq i_o$), the limit of $f(b_1, \ldots, b_n)$ as $b_{i_o} \to -\infty$ exists, and equals zero.

Much deeper and more important are the converse results, giving conditions on f that make it a distribution function. We need the following concept.

Definition: $f: A^n \to$ reals is continuous from below iff for any n-tuple of real numbers, $(x_1, \ldots, x_n)$, and any real $\varepsilon > 0$, there is a real $\delta > 0$ such that $|f(y_1,\ldots,y_n)-f(x_1,\ldots,x_n)| < \varepsilon$ for any $(y_1,\ldots,y_n)$ satisfying $x_i \geq y_i \geq x_i - \delta$, for all $i = 1,\ldots,n$.

Theorem: Let f be a real-valued function with domain A^n. If

(i) f is continuous from below; and

(ii) for every n pairs of real numbers (a_i, b_i), $i = 1 \ldots n$, with $a_i < b_i$, the expression (40) is non-negative, then there is exactly one measure μ such that f is a distribution function for μ in the wide sense.

By combining the last two theorems we get a sufficient condition for f to be a distribution function in the narrow sense — namely (i) and (ii) of the preceding theorem, together with (iii) $f(b_1, \ldots, b_n) \to 0$ as $b_{i_o} \to -\infty$, for any $i_o = 1, \ldots, n$, the other b_i's being held fixed. (Actually, conditions (i), (ii) and (iii) are also necessary for f to be a distribution function in the narrow sense).

Signed Measures

In section 5 we discussed the measure μ_1 on the measurable space $(R \times S \times T, \Sigma_r \times \Sigma_s \times \Sigma_t)$, representing "production" or "births", and also the measure μ_2 on the same space, representing "consumption" or "deaths". One wonders if there is any way of representing "net production", or "natural increase", the difference between these two measures.

Such "netting out" procedures are very common in practice. Thus one subtracts imports from exports to get net exports, in-migration from out-migration to get net migration, debts from credits to get net creditor position, etc.

Formally, one has two measures, say μ and ν, over the same measurable space $(\underline{A},\Sigma)$, and one wants to attach a meaning to the subtraction operation $\mu-\nu$. (In the examples mentioned — exports, migration, and debts — the universe set may be taken to be Space, $\underline{S}$, or perhaps, if one has full "origin-destination" data, $\underline{S} \times \underline{S}$).

The obvious way to define $\mu - \nu$ is as the set-function, with domain Σ, whose value at $\underline{E} \in \Sigma$ is equal to $\mu(\underline{E}) - \nu(\underline{E})$. There are two difficulties with this procedure. First, $\mu - \nu$ will in general take on negative values, and is therefore no longer a measure. Second, if μ and ν are both infinite measures, the meaningless expression $\infty - \infty$ is indicated as the value of $\mu - \nu$ for certain sets (e.g. for the universe set $\underline{A}$ itself); thus things are not even well-defined in this case.

We shall avoid the second difficulty, for the time being, by assuming that at least one of the two measures, μ, ν, is finite. $\mu - \nu$ as defined above is then a well-defined set function on $(\underline{A},\Sigma)$. The important point is that this set function has all of the defining characteristics of a measure, with the single exception that it can take on negative values. This suggests the following definition.

Definition: The set function μ is a signed measure iff

(i) its domain is a sigma-field, and

(ii) its range lies in the extended real numbers, and

(iii) it is countably additive, and

(iv) $\mu(\emptyset) = 0$.

(iii) is the only property that needs explaining. Let $\mathcal{G}$ be a countable packing of measurable sets, and let G_1, G_2, G_3, ... be an enumeration of the members of $\mathcal{G}$; then we must have

$$\mu(\cup\mathcal{G}) = \mu(G_1) + \mu(G_2) + \ldots \qquad (42)$$

in the sense that the limit of the right-hand series exists, and equals $\mu(\cup\mathcal{G})$. Furthermore, this equality is required to hold no matter how the terms of the right-hand series are re-arranged. For measures, where all terms are non-negative, this imposes no additional restriction, since the sum of an infinite series of non-negative terms is invariant under rearrangement of terms. But it <u>is</u> an additional restriction when terms of opposite sign occur in (42). The total requirement may be cast in the following convenient form.

Consider just the <u>positive</u> terms among the $\mu(G_n)$; let the sum of these terms alone be P (as just mentioned, P does not depend on the order of arrangement of these terms; if there are no positive terms, set $P = 0$). Similarly, consider just the <u>negative</u> terms among the $\mu(G_n)$, and let their sum be N (if there are no negative terms, set $N = 0$). Then it is required that, first, at least one of P, N be <u>finite</u>; second, that $P + N = \mu(\cup\mathcal{G})$.

We consider some examples of signed measures:

(i) Any measure is a signed measure; this follows at once from the definitions. (And of course any non-negative signed measure is a measure.)

(ii) Let Σ be a finite sigma-field, generated by partition $\mathcal{G}$, where $\emptyset \notin \mathcal{G}$; assign extended real numbers arbitrarily to the members of $\mathcal{G}$, with the restriction that at most one of the two numbers $\{+\infty, -\infty\}$ is used; for every $\mathcal{F} \subseteq \mathcal{G}$, assign to the set $\cup\mathcal{F}$ the sum of the numbers assigned to the elements of $\mathcal{F}$; finally assign 0 to the empty set $\emptyset$. The result is a signed measure — in fact, all signed measures defined on finite sigma-fields are of this form.

It is trivial that, if a measure is finite above (that is, if it does not take on the value $+\infty$), then it is bounded above. The same property holds for signed measures in general (the proof in this case requiring some effort). In fact we have:

Theorem: If a signed measure is finite above (finite below), then it is bounded above (bounded below).

Thus finiteness and boundedness are synonymous properties for signed measures. From this theorem we obtain the following important property:

Theorem: A signed measure is either bounded above or bounded below (or both).

Proof: Let μ be a signed measure. We show that it cannot take on both values $+\infty$ and $-\infty$. For suppose $\mu(E) = \infty$, $\mu(F) = -\infty$ for some measurable sets E, F; since $\mu(E \cap F) + \mu(F \setminus E) = \mu(F)$, we must have $\mu(E \cap F) \neq \infty$; then, since $\mu(E \cap F) + \mu(E \setminus F) = \mu(E)$, it follows that $\mu(E\setminus F) = \infty$; $E\setminus F$ and F are disjoint, but $\mu(E\setminus F) + \mu(F)$ is undefined, contradicting countable additivity.

Thus a signed measure is finite above, or below, or both; by the theorem above, it is then bounded above, or below, or both. ||||

Definition: Let μ, ν be two signed measures over the same measurable space (A,Σ), such that $\mu(A)$, $\nu(A)$ are not infinite of opposite sign. The sum, $\mu + \nu$, is the set function with domain Σ whose value at E is given by $\mu(E) + \nu(E)$.

(The restriction on $\mu(A)$, $\nu(A)$ ensures that the meaningless expression $\infty - \infty$ does not arise.)

Definition: Let μ be a signed measure over (A,Σ), and let c be a real number. The scalar product, $c\mu$, is the set function with domain Σ whose value at E is given by $c\cdot\mu(E)$.

In particular, $(-1)\mu$ is written simply as $-\mu$.

Theorem: The sum, $\mu + \nu$, and the scalar product, $c\mu$, are signed measures (where μ, ν, and c are restricted as indicated in their respective definitions).

This theorem contains the statement made above concerning the difference of two measures, for $\mu - \nu$ is the same as $\mu + (-\nu)$, and is therefore a signed measure. Again, if μ and ν are measures, then $\mu + \nu$ is always well-defined, and is a signed measure — in fact, a measure, since $\mu + \nu$ is obviously nonnegative.

Let $\mathcal{M}$ be the set of all finite signed measures over space (A,Σ). $\mathcal{M}$ is closed under addition and scalar multiplication. That is, if $\mu \in \mathcal{M}$ and $\nu \in \mathcal{M}$ and c is real, then $\mu + \nu \in \mathcal{M}$, and $c\mu \in \mathcal{M}$. We make the important observation that $\mathcal{M}$ with these two operations is a vector space. Let us define this abstractly.

Definition: Let $\mathcal{M}$ be a set, "+" a function with domain $\mathcal{M} \times \mathcal{M}$ and range in $\mathcal{M}$, and "$\cdot$" a function with domain: (real numbers $\times\ \mathcal{M}$), and range in $\mathcal{M}$. (These are called addition and scalar multiplication, respectively; we use the notation "$\mu + \nu$" and "$c\mu$" instead of the clumsy "$+(\mu,\nu)$" and "$\cdot(c,\mu)$". respectively.) Then $\mathcal{M}$ with these two operations is a (real) vector space iff the following eight conditions are fulfilled:

(i) $(\lambda + \mu) + \nu = \lambda + (\mu + \nu)$, for all $\lambda, \mu, \nu \in \mathcal{M}$;

(ii) $\mu + \nu = \nu + \mu$, for all $\mu, \nu \in \mathcal{M}$;

(iii) there is an element of $\mathcal{M}$, denoted "0", such that $\mu + 0 = \mu$, for all $\mu \in \mathcal{M}$;

(iv) for all $\mu \in \mathcal{M}$, there is an element $\nu \in \mathcal{M}$ such that $\mu + \nu = 0$;

(v) $b(\mu + \nu) = b\mu + b\nu$, for all real b, all $\mu, \nu \in \mathcal{M}$;

(vi) $(b + c)\mu = b\mu + c\mu$, for all real b, c, all $\mu \in \mathcal{M}$;

(vii) $b(c\mu) = (bc)\mu$, for all real b, c, all $\mu \in \mathcal{M}$;

(viii) $1\cdot\mu = \mu$, for all $\mu \in \mathcal{M}$.

The element 0 in (iii) turns out to be unique, and is known as the zero or neutral element of $\mathfrak{M}$. The ν in (iv) is also uniquely determined by μ, and in fact it is equal to $(-1)\mu = -\mu$.

It is clear that this definition has nothing in particular to do with measures, although we have kept the measure notation. The most familiar example of a vector space is the set of all n-tuples of real numbers under the operations "+" and "·" given by

$$(x_1,\ldots,x_n) + (y_1,\ldots,y_n) = (x_1 + y_1,\ldots,x_n + y_n), \text{ and}$$

$$c(x_1,\ldots,x_n) = (cx_1,\ldots,cx_n).$$

We now state formally:

Theorem: Let (A,Σ) be a measurable space, and $\mathfrak{M}$ the set of all finite signed measures on it. Then $\mathfrak{M}$, with the operations $\mu + \nu$ and $c\mu$ defined above, is a vector space.

Proof: We have already stated that $\mathfrak{M}$ is closed under these two operations. The zero element 0 is simply the identically zero measure; the element ν in (iv) is $-\mu$. (i)-(viii) are then immediate consequences of the definitions. |||

Let us now turn to the difficulties raised by infinite measures, signed or unsigned. Throughout this discussion we have had to make qualifications to avoid the meaningless expression $\infty - \infty$. Any attempt to make a vector space out of a larger set of signed measures than M seems to fail, because there is no signed measure ν satisfying

$$\mu + \nu = 0$$

if μ takes on infinite values, so that condition (iv) of the definition of vector space breaks down.

This is unfortunate, because many situations of theoretical interest appear to call for a concept which is, in effect, the difference of two infinite measures. Take net production, which is the signed measure over ($\underline{R} \times \underline{S} \times \underline{T}$, $\Sigma_{\underline{r}} \times \Sigma_{\underline{s}} \times \Sigma_{\underline{t}}$) obtained from the difference, production minus consumption, as discussed above. There is no a priori reason why the two measures, production and consumption, should not both be infinite. Indeed, in problems with an unlimited time-horizon, or an "endless plane", the presumption is that both will be infinite.

Again, suppose one is evaluating economic development policies by comparing costs and benefits against some benchmark. Benefits and costs can be represented as measures on the Time axis. What if both are infinite, a not implausible situation if the horizon is unlimited? One would still like to evaluate benefits minus costs, and if possible to compare two such evaluations.[36]

We have developed the concept of pseudo-measure to overcome these difficulties. This concept, being outside the corpus of present-day measure theory, deserves full-scale treatment in a chapter of its own, so we shall not define it here. But we do wish to indicate how it jibes with the more familiar concepts presently under discussion. Briefly, pseudomeasures are generalizations of sigma-finite signed measures, just as the latter are generalizations of sigma-finite measures. With their aid sigma-finite signed measures can be added freely, even when infinite of opposite sign. When one extends $\mathfrak{M}$ to all sigma-finite signed measures, it ceases to be a vector space; but extending it still further to all pseudomeasures restores this property. Furthermore one can order pseudomeasures in ways that are elegant and intuitively appealing with respect to the problems mentioned above.

The present discussion of signed measures may be viewed as a halfway point, to be suitably generalized when we come to pseudomeasures in chapter 3.

We have seen that, for any pair of measures μ, ν, not both infinite, the difference $\mu - \nu$ is a signed measure. A basic result is that the converse is also true: Any signed measure can be expressed as the difference of two measures.

Theorem: (Jordan decomposition theorem) Let μ be a signed measure on space (A,Σ), and consider the following two set-functions, both with domain Σ.

$$\mu^{+}(E) = \sup\{\mu(F) \mid F \subseteq E,\ F \in \Sigma\}, \tag{43}$$

$$\mu^{-}(E) = \sup\{-\mu(F) \mid F \subseteq E,\ F \in \Sigma\}. \tag{44}$$

Then

μ^+ and μ^- are measures, not both infinite, and $\mu = \mu^+ - \mu^-$. (In (43) and (44), "sup" abbreviates supremum, which is taken over the set of all values assumed by μ respectively, $-\mu$ on measurable subsets of E).

Definitions: μ^+ in (43) is known as the upper variation of μ, μ^- in (44) as the lower variation of μ. The sum $\mu^+ + \mu^-$ is known as the total variation of μ, and is denoted $|\mu|$. The pair (μ^+, μ^-) is the Jordan decomposition of μ.

For example, let μ be a measure; then $\mu^+ = \mu$, and $\mu^- = 0$. (Proof: by monotonicity, $\mu(E) \geq \mu(F)$ if $F \subseteq E$; hence $\mu^+(E) = \mu(E)$; since μ is non-negative, the supremum of $-\mu$ is attained on the empty set $\emptyset$; hence $\mu^-(E) = 0$). Similarly, if μ is non-positive (that is, μ is the negative of a measure), then $\mu^+ = 0$, and $\mu^- = -\mu$.

Suppose one starts with a pair of measures, (μ_1, μ_2) (not both infinite), forms their difference $\mu = \mu_1 - \mu_2$, and then takes the Jordan decomposition (μ^+, μ^-). What is the relation between these two pairs? The answer is given by the following.

Theorem: Let μ_1, μ_2 be two measures over (A, Σ), not both infinite, and let $\mu_1 - \mu_2 = \mu$. Then there is a finite measure ν such that

$$\mu_1 = \mu^+ + \nu, \text{ and } \mu_2 = \mu^- + \nu, \qquad (45)$$

Proof: If E, $F \in \Sigma$, and $F \subseteq E$, then $\mu(F) \leq \mu_1(F) \leq \mu_1(E)$. It follows from (43) that $\mu^+(E) \leq \mu_1(E)$, for all $E \in \Sigma$. Similarly, $-\mu(F) \leq \mu_2(F) \leq \mu_2(E)$, so $\mu^-(E) \leq \mu_2(E)$, for all $E \in \Sigma$, from (44).

If μ_2 is finite, so is μ^-; in this case set $\mu_2 - \mu^- = \nu$; ν is a finite measure, and the relation $\mu_1 - \mu_2 = \mu^+ - \mu^-$ yields (45).

If μ_1 is finite, so is μ^+; set $\mu_1 - \mu^+ = \nu$, and we again get (45). □

Thus for given signed measure μ, the Jordan decomposition is the <u>smallest</u> pair of measures whose difference is μ. Any other pair of measures having this property is obtained by adding the same finite measure ν to both halves of (μ^+, μ^-). The Jordan decomposition also has the following deeper property.

Definition: Measures μ, ν over space (A, Σ) are <u>mutually singular</u> iff there is a partition of A into two measurable sets, P, N (i.e., $P \cup N = A$, $P \cap N = \emptyset$) such that $\mu(N) = 0$ and $\nu(P) = 0$.

Theorem: (Hahn decomposition theorem) The upper and lower variations, μ^+ and μ^-, of any signed measure μ are mutually singular.

This may be stated in the following slightly different form.

Theorem: (Hahn decomposition theorem, second version) For any signed measure μ on space (A,Σ), A can be split into two measurable sets, P, N such that, for all $E \in \Sigma$,

(i) if $E \subseteq P$, then $\mu(E) \geq 0$; and

(ii) if $E \subseteq N$, then $\mu(E) \leq 0$. (46)

We shall prove that these two versions imply each other. Assume the first, and let P, N be a measurable partition of A, such that $\mu^+(N) = 0$ and $\mu^-(P) = 0$. If measurable $E \subseteq P$, then $\mu(E) = \mu^+(E) - \mu^-(E) = \mu^+(E) \geq 0$. If $E \subseteq N$, then $\mu(E) = \mu^+(E) - \mu^-(E) = -\mu^-(E) \leq 0$. This proves version two.

Conversely, assume the second version. $\mu(E) \leq 0$ for every measurable subset of N; hence $\mu^+(N) = 0$, by (43). Similarly, $-\mu(E) \leq 0$ for $E \subseteq P$, $E \in \Sigma$; hence $\mu^-(P) = 0$. Thus μ^+, μ^- are mutually singular. ///

Definition: Given signed measure μ on (A,Σ). Any ordered measurable partition (P,N) of A satisfying (46) will be called a Hahn decomposition for μ. P is the <u>positive half</u>, N the <u>negative half</u>, of this decomposition.

Equivalently, this could have been defined as a measurable partition satisfying $\mu^-(P) = 0$, $\mu^+(N) = 0$.

Thus while the Jordan decomposition is a pair of measures, the Hahn decomposition is a pair of sets. The Jordan decomposition is unique; the Hahn decomposition is "almost" unique in the sense that if (P',N') is another Hahn decomposition, then $|\mu|(P \cap N') = 0$ and $|\mu|(P' \cap N) = 0$, $|\mu|$ being the total variation of μ. Given P, $\mu^+(E) = \mu(E \cap P)$, since, within E, μ takes on its supremum at $E \cap P$; similarly, $\mu^-(E) = -\mu(E \cap N)$.

Among all pairs of measures whose difference is μ, the Jordan decomposition is the only pair which is mutually singular.

As an example, let μ be a measure. Then $P = A$, $N = \emptyset$ is a Hahn decomposition for μ. (Proof: immediate from (46)).

The Hahn and Jordan decompositions of real-world signed measures have simple intuitive interpretations. Suppose, for example, that the universe set is Space, S, and let signed measure μ be net exports of some commodity: $\mu(E)$ = exports from E minus imports to E, for every region E. The Hahn decomposition theorem shows that S can be split into two regions, P and N, such that every subregion of P (N) is a net exporter (importer). μ^+ and μ^- are then the export and import measures exclusive of transhipment.[37]

Most of the theorems we have discussed have generalizations from measures to signed measures. Because of the Hahn and Jordan decomposition theorems, it is unnecessary for the most part to state these separately: One simply performs the appropriate decomposition, applies the theorem in question to each piece separately, and then puts the pieces together to get the generalization to signed measures. There are just two cautions to be observed: (i) Theorems involving inequalities do not always generalize; thus the simple statement $\int_A f\, d\mu \geq 0$ need not hold if f or μ can assume negative values (see below for definition). (ii) One sometimes needs additional assumptions to guarantee that the meaningless expression $\infty - \infty$ does not arise.[38]

There are a few concepts whose generalization to signed measures deserves explicit mention. We have already mentioned sigma-finiteness, whose definition carries over without change.

Definition: Let μ be a signed measure on (A,Σ). μ is *sigma-finite* iff there is a countable partition, $\mathcal{G}$, of A into measurable sets, such that $\mu(G)$ is finite for all $G \in \mathcal{G}$.

We now come to integration. Given measurable space (A,Σ), the integral $\int_A f\, d\mu$ has been defined in the case where μ is a measure and f a non-negative measurable function. We shall now remove both of these sign restrictions.

Definition: Let f be an extended real-valued function on domain A. f^+ and f^- are functions on domain A given by

$$f^+(a) = \max(f(a),0), \quad f^-(a) = \max(-f(a),0).$$

("max" abbreviates "maximum"; that is, f^+ coincides with f where the latter is positive, and equals zero elsewhere; f^- coincides with $-f$ where f is negative, and equals zero elsewhere). f^+ is known as the *positive part* of f, f^- as the *negative part* of f.

Note that both f^+ and f^- are non-negative functions. Also that $f^+ - f^- = f$, while $f^+ + f^- = |f|$, the absolute value of f.

Definition: Let μ be a signed measure, and f a measurable function on the measurable space (A,Σ). The integral of f with respect to μ is given by:

$$\int_A f\, d\mu = \left(\int_A f^+\, d\mu^+\right) + \left(\int_A f^-\, d\mu^-\right) - \left(\int_A f^+\, d\mu^-\right) - \left(\int_A f^-\, d\mu^+\right), \quad (47)$$

provided the right-hand expression is not of the form $\infty - \infty$. (If it is of this form, then $\int_A f\, d\mu$ is considered to be meaningless.)

(In each of the four integrals on the right of (47), the integrands, f^+ and f^-, are non-negative, and μ^+ and μ^- ~~which are of course~~ (the upper and lower variations of μ) are measures. Thus these integrals have already been defined, and (47) gives the integral for signed measures and "signed functions" in terms of these already defined integrals). 39

Definition (47) is consistent in the following sense: If f is non-negative, and μ is a measure, then the definition in (47) coincides with the original. This follows from the fact that, in this case, $f^+ = f$, $f^- = 0$, $\mu^+ = \mu$, $\mu^- = 0$; thus three of the four integrals are zero, and the last gives

$$\int_A f^+\, d\mu^+ = \int_A f\, d\mu,$$

where the right-hand expression has the original definition, (34).

Two "half-way" cases arise. If μ is a measure, (47) reduces to

$$\int_A f^+ \, d\mu - \int_A f^- \, d\mu,$$

and, if f is non-negative, (47) reduces to

$$\int_A f \, d\mu^+ - \int_A f \, d\mu^-.$$

As an example, let $f(a) = c$ for some fixed real number c. Then (47) yields:

$$\int_A c \, d\mu = c\mu(A),$$

exactly the same formula as when $c \geq 0$ and μ is a measure. (Proof: consider separately the two cases, $c \geq 0$, $c < 0$).

We conclude this discussion with a few integration formulas. On measurable space (A,Σ), μ and ν are signed measures, f and g measurable extended real-valued functions, c is a real number.

If f and μ are both bounded, then

$$\int_A f \, d\mu \qquad (48)$$

is well-defined, and is finite.

$$c \int_A f \, d\mu = \int_A cf \, d\mu. \tag{49}$$

$$c \int_A f \, d\mu = \int_A f \, d(c\mu). \tag{50}$$

$$\int_A f \, d\mu + \int_A g \, d\mu = \int_A (f + g) d\mu \; . \tag{51}$$

$$\int_A f \, d\mu + \int_A f \, d\nu = \int_A f \, d(\mu + \nu). \tag{52}$$

(49) through (52) are to be read as follows: If both sides are well-defined, then they are equal. (48), (49), and (51) thus generalize (9), (10), and (11) respectively. In (50) $c\mu$ is the scalar product; in (52) $\mu + \nu$ is the sum of two signed measures.

2.7. Activities

The measure space of histories (Ω, Σ, μ) — which in principle gives a complete description of the world — is rather unwieldy as a whole, and one wants to focus attention on one or another aspect of special interest. We have already discussed how certain data may be extracted — cross-sectional and double cross-sectional measures, production and consumption, and the like. Here we continue this discussion, concentrating not so much on material which typically appears in statistical tabulations, but on less sharply defined categories: "situations," "events", "processes", "activities".

Thus out of the flux one distinguishes a house, a crowd of people, a town, a seacoast; or, a driver traveling along the highway, a sugar refinery in operation, a farmer plowing his field, an army on the march. The first four items mentioned are "cross-sectional configurations"; that is, they describe a part of the world at an instant of time. The last four refer to something going on over an interval of time.

One can distinguish situations in an indefinitely large variety of ways. Out of these possibilities a much smaller -- but still enormous -- number are actually distinguished and named in the words of some language. Why some possibilities are selected and others not is itself an interesting question, to be answered on the one hand by relations of similarity, contiguity, contrast, closure and other characteristics of "good gestalt", and on the other by causal relations. Situations tend to be selected so that their parts are mutually inter-dependent, and relatively independent of the rest of the world.

We shall have something to say about causation in chapter 4. In the present chapter, however, we are concerned only with problems of description. What we want, then, is a frame-work adequate for describing a variety of situations or processes, whether they make "causal sense" or not. If one takes the stork population of Sweden and the human birth rate of that country as constituents of a single activity, this activity can be perfectly well defined (although not very useful perhaps).

Cross-sectional configurations have ~~already~~ been discussed briefly in Section 2.5 ~~(p.)~~. It was mentioned that a house, for example, could be represented as a measure over universe set R × F, R being the set of resource-types, and F the region occupied by the house. This measure gives the spatial distribution of the resources constituting the house. ~~There~~ was one point left hanging ~~, however,~~ in this discussion. A measure over R × F represents only a particular house, viz., ~~— namely~~ the house occupying the particular region F (at the moment for which the cross-section was defined). We have no way as yet for representing a house-*type*, or configuration-type in general, as opposed to any particular specimen of that type. The following construction fills this gap.

Metric Spaces and Congruent Measures

Definition: A *metric space* consists of a set, A, and a real-valued function, d, with domain A × A, satisfying

(i) d(x,x) = 0, for all x ∈ A; ~~and~~

(ii) d(x,y) > 0 if x ≠ y, for all x, y ∈ A; ~~and~~

(iii) d(x,y) = d(y,x), for all x, y ∈ A; ~~and~~

(iv) d(x,y) + d(y,z) ≥ d(x,z), for all x, y, z ∈ A.

(Condition (iii) is called symmetry, condition (iv) the triangle inequality. The metric space itself is written as the pair (A,d). d is called the metric, or the distance function. If d is understood, A itself may be referred to as a metric space).[40]

Our first example is the most familiar case.

Definition: Let A be n-space. The Euclidean metric gives the distance between $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$ as

$$d(x, y) = [(x_1 - y_1)^2 +\ldots+ (x_n - y_n)^2]^{1/2}. \qquad (1)$$

Conditions (i), (ii), and (iii) on d are verified immediately. (iv), which is a little harder to verify, states exactly that, in the triangle with vertices x, y, and z, the length of the side from x to z does not exceed the sum of the lengths of the other two sides.

Definition: Let A again be n-space. The city-block metric (also known variously as the rectangular, metropolitan, manhattan, or midwestern metric) gives the distance between $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$ as

$$d(x,y) = |x_1 - y_1| +\ldots+ |x_n - y_n| \qquad (2)$$

Conditions (i) through (iv) on the d defined by (2) are immediate consequences of the properties of absolute values.

When n = 1 we have the real line, and in this case (1) and (2) both reduce to the same function, namely,

$$d(x, y) = |x - y| .$$

But for $n > 1$ the two metrics are distinct. The most important case for our later work will be $n = 2$, the resulting metric spaces being called the Euclidean plane and the city-block plane, respectively.

As a third example, let A be the surface of a sphere of radius r and center $c = (c_1, c_2, c_3)$. The great-circle metric gives the distance between $x, y \in A$ as $r \cdot$ angle $(x c y)$, the angle being measured in radians.

~~One~~ As a final example. Let A be any non-empty set, and define d by: $d(x, y) = 1$ if $x \neq y$; $d(x, x) = 0$. Since this satisfies (i) ~~through~~ (iv) it is a bona fide distance function, known as the discrete metric.

Let (A,d) be a metric space, and B a subset of A. B can be considered a metric space in its own right by taking $d(x, y)$, for $x, y \in B$, to be the distance from x to y in metric space B. This amounts to defining the metric on B to be the restriction of d to the sub-domain $(B \times B) \subseteq (A \times A)$. We ~~shall~~ always consider any subset of a metric space to be itself a metric space in this way.

Definition: Let (A,d), (B,d') be two metric spaces. A mapping $f: A \to B$ is a congruence, or an isometry, from A to B, iff

(i) f is onto (~~that is~~ i.e., for all $b \in B$, ~~there is~~ an $a \in A$ exists for which $f(a) = b$); ~~and~~

(ii) $d(x, y) = d'(f(x), f(y))$, for all $x, y \in A$. (2.7.3) ~~(3)~~

These two metrics are examples of normed metrics; ~~That~~ i.e. is, A is a vector space, and there exists a function, $\|\cdot\|$, on A (called a norm) satisfying: $\|x\| > 0$ for all $x \neq 0$, $\|x + y\| \leq \|x\| + \|y\|$, $\|px\| = |p| \cdot \|x\|$ for p real, and such that $d(x, y) = \|x - y\|$, all $x, y \in A$

(3) states that f is distance-preserving. A and B may be overlapping, or even identical, and, in the latter case, d' may or may not be the same as d.

Definition: Metric spaces (A,d) and (B,d') are *congruent*, or *isometric*, iff there exists an isometry $f: A \to B$.

If (A,d) is congruent to (B,d'), which in turn is congruent to (C,d''), then (A,d) is congruent to (C,d''). This follows from the fact that, if $f: A \to B$ and $g: B \to C$ are isometries, then the composition $g \circ f: A \to C$ is an isometry. Thus the relation of congruence between metric spaces is *transitive*. Furthermore, if (A,d) is congruent to (B,d'), then (B,d') is congruent to (A,d). For, if $f: A \to B$ is an isometry, then it has an *inverse* function $g: B \to A$ (that is, $g(f(a)) = a$, all $a \in A$, and $f(g(b)) = b$, all $b \in B$), and g is also an isometry. Thus congruence is a symmetric relation. Finally, (A,d) is obviously congruent to itself, the identity map $f(a) = a$ being an isometry; thus congruence is *reflexive*. In short, congruence is an *equivalence* relation between metric spaces.

Next, let (A,Σ) and (B,Σ') be two measurable spaces.

Definition: $f: A \to B$ is *measurability-preserving* (from A to B) iff

(i) f has an inverse function $g: B \to A$, and

(ii) both f and g are measurable.

(Measurability of f means of course that $\{a|f(a) \in E\}$ belongs to Σ whenever $E \in \Sigma'$; measurability of g reverses these roles, so that $\{b|g(b) \in F\}$ belongs to Σ' whenever $F \in \Sigma$. Condition (i) holds iff f is 1-1 and onto).

Let (A,Σ,μ) and (B,Σ',μ') be two measure spaces.

Definition: $f: A \to B$ is measure-preserving iff

(i) f is measurability-preserving, and

(ii) $\mu'(E) = \mu\{a|f(a) \in E\}$, for all $E \in \Sigma'$.

(Condition (ii) is of course just the statement that μ' is the measure induced by f from μ. This is all completely symmetric, because μ is also the measure induced from μ' by the inverse mapping g.)

These concepts may now be combined. Suppose A is provided with a metric, d, and with a sigma-field, Σ, and measure, μ; thus A is both a metric space and a measure space, and we write this as a quadruple (A,d,Σ,μ).[41]

Definition: Let (A,d,Σ,μ) and (B,d',Σ',μ') be two metric-measure spaces. These are measure-congruent iff there exists an isometry $f: A \to B$ (for the pair (A,d) and (B,d')) which is also measure-preserving (for the pair (A,Σ,μ) and (B,Σ',μ')).

(Note that there must be a single function which simultaneously preserves distance and measure. If the metrics are understood, one says that μ and μ' are congruent measures.)

To illustrate these concepts let us take the Euclidean plane, with the two-dimensional Borel field and Lebesgue measure. Let E, F be two measurable sets which are congruent in the sense of plane geometry: they have the same "size" and "shape". That is, there is a function $f: E \to F$, whose range is all of F, and for which $d(x,y) = d(f(x), f(y))$, all $x, y \in E$, where d is Euclidean distance. Letting Σ_E, Σ_F be the Borel field restricted to subsets of E, F, respectively, one can show that f is measurability-preserving. In fact, f is measure-preserving with respect to Lebesgue measure restricted to (E,Σ_E) and (F,Σ_F), respectively. Hence E and F are measure-congruent.

Measure-congruence is an equivalence relation among metric-measure spaces, just as congruence per se is an equivalence relation among metric spaces.

Configuration Types

Let us now apply these concepts. We suppose that Space, S, has a metric d and a sigma-field Σ (e.g. S is 3-space, d the Euclidean metric, Σ the Borel field). Let E and F be two regions. As discussed above, these may be thought of as metric spaces, and also as measurable spaces, in their own right.

Let μ, ν be two distributions of mass over universe sets E, F, respectively. (We are suppressing discussion of the Resource set R for the moment; μ and ν may be thought of as the marginals of measures over $R \times E$ and $R \times F$, respectively). It intuitively appealing to explicate the vague relation, "μ and ν the same type of configuration" by the precise relation, "μ and ν are measure-congruent". The latter requires that E and F have the same size and shape, and that the distributions over E and F have the same "pattern" — so that ν is in a sense just a rigid displacement of μ.[42]

We now bring Resources, R, into the discussion. Let μ, ν be mass distributions over spaces $(R \times E, \Sigma_r \times \Sigma_E)$ and $(R \times F, \Sigma_r \times \Sigma_F)$, respectively. Here E and F are regions, as above; Σ_E and Σ_F are the restrictions of the sigma-field Σ_s of S to measurable subsets of E, F, respectively; and Σ_r is the sigma-field of R; regions E and F are provided with the metric of Space, but R is not assumed to have any metric.

Definition: We say that μ and ν represent the same configuration-type iff

(i) there is an isometry $f: E \to F$ which is also measurability-preserving [with respect to (E, Σ_E) and (F, Σ_F)]; and

(ii) for all $G \in \Sigma_r \times \Sigma_F$,

$$\nu(G) = \mu\{(r,s) \mid (r, f(s)) \in G\}. \tag{4}$$

Note that f is a mapping between regions, and is not involved directly with R at all. However, f determines a certain mapping from $R \times E$ onto $R \times F$, namely, the one carrying the point (r,s) to $(r, f(s))$: Location shifts, but resource-type is held fixed. (4) then asserts that ν is the measure induced by this mapping from μ.

(Indeed, it ~~is easy to~~ see that the mapping $(r,s) \to (r, f(s))$ is measure-preserving. First ~~of all~~, it is measurable, since the two functions $(r,s) \to r$ and $(r,s) \to f(s)$ are both measurable. Second, since f has an inverse function g, the mapping $(r,s) \to (r, g(s))$ from $R \times F$ onto $R \times E$, is the inverse of $(r,s) \to (r, f(s))$; and it is measurable since g is).

If μ and ν represent two houses, say, then condition (i) requires that they have the same size and shape, while condition (ii) requires that the same materials (wood, glass, brick, ~~etc.~~) be arranged in the same relative positions in both houses. Thus this definition appears to capture quite well the intuitive notion of "two instances of the same kind of thing".

Resources and Space are not treated symmetrically in this definition. The reason lies in the "heterogeneous" nature of R as opposed to the "homogeneous" nature of S. There is no analog in R to the congruence relation among regions of Space — at least not in general.[43]

If μ and ν represent the same configuration-type, then their right-marginals, μ' and ν', on spaces (E,Σ_E) and (F,Σ_F), respectively, are congruent measures. This follows from (4) upon taking $G = R \times F'$, where $F' \in \Sigma_F$:

$$\nu'(F') = \nu(R \times F') = \mu\{(r,s)|f(s) \in F'\} = \mu'\{s|f(s) \in F'\}.$$

We have defined the condition under which two specific configurations are to be considered of the same type. But we have not yet defined the notion of configuration-type abstractly, divorced from any particular specimen of that type. For example, we can say that this house in Lisbon is of the same type as that one in Hong Kong, but cannot yet speak of this house-type not in any particular place. Recall that we conceive of Space, S, in real terms, its regions being actual geographical places, so that our definition above is "tied" to the real world.

This difficulty is easily overcome. We suppose R and its sigma-field Σ_r to be given. A configuration-type is then defined as a quadruple, consisting of a set, E, a metric, d, on E, a sigma-field, Σ_E, on E, and a measure, μ, on ($R \times E$, $\Sigma_r \times \Sigma_E$). This configuration-type is exemplified on region F at moment t iff there is a measurability-preserving isometry, $f: E \rightarrow F$, such that the mass-distribution, ν, over $R \times F$ at time t, satisfies (4) with μ.

In other words, the definition of "exemplification" is formally identical to that of "same configuration-type", and merely differs in interpretation. In the latter, E and F are both regions, and we get a relation between two particulars; in the former, E is an abstract set endowed with a metric and sigma-field, and we get a relation between a universal and a particular. Furthermore, one can run this definition in reverse, and identify the abstract configuration-type of any particular: merely strip the region it occupies of all properties except its metric and sigma-field.

There is no need to abstract from R as we did from S, because R is already the set of *types* of resources.

One comment on the relation between these concepts and ordinary language. It is clear that "house", "forest", "crowd of people" and the like refer not to a single configuration-type, but instead roughly delimit a set of such types. The more detailed the description, the smaller the set of configurations satisfying it; but even an encyclopedic description does not narrow the range down to one. (Furthermore -- and logically distinct from the ambiguity just mentioned -- there is a penumbra of vagueness about ordinary language, so that there are "borderline cases" and "twilight zones" where one is not sure whether a given configuration-type satisfies a given description or not.)

Suppose one has a description of the world at some moment, *t*, in the form of a list of configuration-types exemplified in various regions, E_1, E_2, ... Some of these regions may overlap: If $F \subseteq (E_i \cap E_j)$ where F is another region, this means that the mass distributed over F participates simultaneously in two configurations. For example, the ceiling of one apartment may be the floor of another. If the descriptions are accurate, the two measures μ_i and μ_j will then be identical over F. We know from the Patching theorem (p.) that this is a sufficient condition for the various μ_i's to determine a unique measure μ on the universe set $\cup\{E_n | n = 1, 2, \ldots\}$, μ_n being the restriction of μ to E_n. Thus the complete cross-sectional measure may be "patched" together.

In particular, if $E_i \subseteq E_j$ we have the relation of part and whole (a neighborhood in a city, a family in a nation, etc.). That is, the relation of whole to part is represented formally by the measure μ_j over $R \times E_j$ and its restriction to $R \times E_i$, which is μ_i. One frequently deals with a whole hierarchy of parts and wholes: items into packages into cartons into car-loads into trainloads, etc.

Activities

Our discussion thus far has been confined to cross-sectional measures. We now turn to "dynamic configurations", so to speak. "Activity" will be used as a generic term for such processes. There seem to be several related but distinct concepts here.

Let us start with the measure space of histories (Ω, Σ, μ). Just as a configuration is a restriction of a cross-sectional measure, one may think of an activity as a restriction of μ itself. That is, one takes a measurable subset of histories, $H \in \Sigma$, and refers to μ restricted to H as an activity.[44] For example, H may be the set of all histories whose itineraries lie in region F, and whose transmutation-paths lie in resource-set E at moment t; more generally, which lie in $E_i \times F_i$ at moment t_i $(i = 1, \ldots, n)$. Or H can be the set of histories lying in $E \times F$ ($\in \Sigma_r \times \Sigma_s$) at least once during time-interval G, or, which lie in $E \times F$ throughout time-interval G.

A different approach starts not with (Ω,Σ,μ) but with the production and consumption measures, λ_1 and λ_2, derived from it. Recall that the universe set here is $R \times S \times T$, with the interpretation: $\lambda_1(E \times F \times G)$ = mass of all histories "produced" or "born" during period G, and starting in resource-set E and region F. λ_2 is the same, with "consumed" or "dying" in place of "produced" or "born". An activity in this sense is a restriction of λ_1 or λ_2 (or both, or net production $\lambda_1 - \lambda_2$) to a subset of $R \times S \times T$ of the form $R \times E$, E a measurable subset of $S \times T$.

It is very common and useful to combine these approaches. Consider, for example, the description of the operation of a certain shoe factory, say from time t_1 to t_2. The appropriate set of histories H consists of all which act as "factors of production" for some time during the interval $[t_1, t_2]$ – the factory building, the land on which it sits, the tools, the machinery, the workers and management personnel – even the air and the gravitational field at the site. For production and consumption, take the restricted set $R \times F \times [t_1, t_2]$, F being the region occupied by the factory and grounds. λ_1 will include the production of shoes, but also the production of smoke, noise, odor, leather scraps, etc. λ_2 will include the consumption of rubber, leather, nails, glue, water, electricity, fuel, etc. The activity in question may now be defined as the triple of measures (μ, λ_1, λ_2), μ restricted to $\underline{H}$, and λ_1, λ_2 to $\underline{R} \times \underline{F} \times [\underline{t}_1, \underline{t}_2]$, or as the pair ($\mu$, $\lambda_1 - \lambda_2$), $\lambda_1 - \lambda_2$ being a signed measure or, more generally, a pseudomeasure.

This is by no means the only possible representation of this process. For one thing, it omits inventories of materials; raw, in-process, and finished. These may be incorporated, if desired, by expanding the set of histories $\underline{H}$ to include them. (The distinction between materials and factors is in any case not a sharp one. Materials change form - (i.e., position in $\underline{R}$) - relatively fast, factors relatively slowly). ~~On the other hand,~~ it is sometimes convenient to treat the commenting workforce as if it were consumed upon arriving each morning, and produced a-new upon leaving. In this representation labor would be recorded in the λ_1, λ_2 accounts, each worker being counted once for each separate commuting trip he makes.

Let us try to classify various activities. A first distinction is between activities ~~which~~ "stay put" and those ~~which~~ change location over time. Letting $\underline{H}$ be the set of histories over which a given activity, $\underline{q}$, is defined, the <u>location</u> of $\underline{q}$ at time $\underline{t}$ is the region

$$\underline{F}_{\underline{t}} = \{\underline{x} | \underline{h}_{\underline{s}}(\underline{t}) = \underline{x} \text{ for some } \underline{h} \in \underline{H}\}. \qquad (5)$$

(Recall that $\underline{h}_{\underline{s}}$ is the itinerary of history $\underline{h}$. The fictitions point $\underline{z}_{\underline{o}}$, indicating that a given history is not in existence at instant $\underline{t}$, is excluded from this set). If $\underline{F}_{\underline{t}}$ is constant over the interval $[\underline{t}_1, \underline{t}_2]$, the activity is said to be sedentary in that interval.

One can also define the location of an activity in terms of the production-consumption measures, λ_1 and λ_2. Their universe set is $\underline{R} \times \underline{E}$, and the spatial cross-section of $\underline{E}$ at time $\underline{t}$; $\{s | (s,t) \in \underline{E}\}$ is the location of the activity at $\underline{t}$. If an activity is defined having both μ (stock) and λ (flow) components, it is generally convenient to make this set coincide with $\underline{F}_t$ of (5).

Is the shoe factory activity discussed above sedentary in the interval $[\underline{t}_1, \underline{t}_2]$? Not according to the original definition, because the workforce commutes in and out of the plant site. But if one transfers labor, and any other factors entering or leaving the plantsite during the interval in question, to the production-consumption account, as suggested above, then it becomes sedentary. This shows that whether a certain process is to be considered sedentary ~~or not~~ is at least partly a matter of convention.

A simply-located activity[45] is one whose location at any instant is a single point of Space (or the empty set). If this single point is the same over the interval $[\underline{t}_1, \underline{t}_2]$ ~~then~~ it is also a sedentary activity. The concept of simple location for an activity is of course an idealization, but a very useful one, as chapter 8 will demonstrate.

Among sedentary activities we distinguish those which are stationary, or steady-state. First let us look at activities defined in terms of production and consumption over universe set $R \times F \times T$, F being a region. T is, as usual, represented by the real numbers. λ, over the measurable space $(R \times F \times T, \Sigma_r \times \Sigma_F \times \Sigma_t)$ is stationary iff, for all $E \in \Sigma_r \times \Sigma_F \times \Sigma_t$, and for all real numbers c,

$$\lambda(E) = \lambda\{(r,s,t) \mid (r,s,t + c) \in E\}. \tag{6}$$

Here λ can stand for production or consumption, or their difference, net production. (6) states that, say, production, is invariant under displacement in time, and this captures the intuitive notion of a "steady rate of production".

The concept of steady-state for activities defined in terms of histories is a bit more complicated.

Definition: Let h be a function whose domain is the real numbers with values in a set B, and let c be a real number; the c-displacement of h is the function h^c: reals $\rightarrow B$ given by

$$h^c(x) = h(x - c), \tag{7}$$

all real x. Interpreting the domain as Time, (7) states that wherever h goes, h^c goes c time units later ($-c$ time units earlier, if c is negative). If h is a history, then h^c is also a history, for all real c.

Now let H be the set of histories over which an activity q, with restricted measure μ, is defined. q is stationary iff

(i) H is closed under displacement. (~~That is~~, if $h \in H$, then the c-displacement of h also belongs to H for all real c); ~~and~~

(ii) if $E \subseteq H$ is measurable, and c is real, then

$$\mu(E) = \mu\{h \mid h^c \in E\}.$$

Condition (i) implies that q is sedentary; condition (ii) that the cross-sectional distribution of mass for the histories H at time t is independent of t, that the double-cross-sectional distribution at times t_1 and t_2 depends only on the difference $t_2 - t_1$, and, in general, that the entire process "looks the same" if shifted arbitrarily through time.

Finally, if an activity is given by a pair $(\mu, \lambda_1 - \lambda_2)$ or triple $(\mu, \lambda_1, \lambda_2)$ it is to be considered stationary iff all its components are stationary according to the respective definitions above, ~~the same number c satisfying all simultaneously.~~

Stationarity is a severe requirement. Under it there can be no batch production, only a continuous flow; no shifts, only a continuous arrival and departure of workers; no daily, weakly, or seasonal cycles. It is the ultimate in uneventfulness.

Activity Types

Just as with configurations, one distinguishes particular activities, located in specific portions of Space-Time, and activity-types. We shall follow the same procedure as above; namely, first to determine when two specific activities are considered to be of the same type, and second to define activity-type abstractly.

Let f be a measurability-preserving isometry of Space onto itself. We shall consider two activities to be of the same type iff there is such an f which, together with a time displacement, transforms one of these activities into the other.

First consider two activities defined in terms of production or consumption: say λ on the universe set $R \times G$, and λ' on the universe set $R \times G'$ (where G and G' are measurable subsets of $S \times T$). λ and λ' are then said to be of the same activity-type iff there is an $f: S \to S$ as above and a real number c such that

$$G = \{(s,t) \mid (f(s), t + c) \in G'\}, \tag{8}$$

and such that

$$\lambda'(E) = \lambda\{(r,s,t) \mid (r, f(s), t + c) \in E\}, \tag{9}$$

for all measurable $E \subseteq R \times G'$.

(8) states that the two Space-Time "regions", G and G' have the same "size" and "shape", while (9) states that the relative patterns of production (or consumption, etc.) within these "regions" is the same. Note that these activities need not be sedentary.

Next take two activities defined in terms of histories: say μ on universe set H, and μ' on H' (H and H' being measurable subsets of Ω). With $f: S \to S$ as above, c a real number, and h a history, the c, f-transformation of h is the history $h^{c,f}$ given by the following rules:

For the transmutation-path:

$$h_r^{c,f}(t) = h_r(t - c), \text{ all } t \in T.$$

For the itinerary:

$$h_s^{c,f}(t) = f(h_s(t - c)), \text{ all } t \in T.$$

(If $h(t - c) = z_o$, it is understood that $h^{c,f}(t) = z_o$.)

That is, the transmutation-path of $h^{c,f}$ is simply the c-displacement of the transmutation-path of h; this is unaffected by f. The itinerary of $h^{c,f}$ is the f-transformation of the c-displacement of the itinerary of h. (If f is the identity, this simply reduces to the c-displacement of h).

μ and μ' are then said to be of the same activity-type iff there is a measurability-preserving isometry $f: S \to S$, and a real number c such that

$$H = \{h \mid h^{c,f} \in H'\}, \qquad (10)$$

and such that

$$\mu'(E) = \mu\{h \mid h^{c,f} \in E\}, \qquad (11)$$

for all measurable $E \subseteq H'$.

Finally, if two activities q and q' are both triples, $(\mu, \lambda_1, \lambda_2)$ and $(\mu', \lambda_1', \lambda_2')$, respectively, or pairs (μ, λ) and (μ', λ'), they are said to be of the same activity-type iff each of the components are of the same activity type according to the above respective definitions, with the same f and c satisfying all simultaneously.

The similarity between these definitions and that of stationarity is patent. In fact, an activity is stationary iff it is of the same type as any time displacement of itself, Space being held fixed.

We now define the concept of activity-type in the abstract, first for activities of the production-consumption type. We are given the measurable space of Resources, (R, Σ_r). Let (T', Σ_t') be the real line and its Borel field. An activity-type is then defined as a quintuple, consisting of a set, S', a metric, d', on S', a sigma-field, Σ_s', on S', a subset G' of $S' \times T'$ which is measurable; $G' \in \Sigma_s' \times \Sigma_t'$; and a measure, λ', on universe set $R \times G'$.

(Here the notation S', T' is meant to suggest "abstract" Space and Time; there is no need to abstract from R since this is already a set of resource-types. λ' then gives the production or consumption pattern over the abstract space $R \times G'$; if we are dealing with net production, then λ' is a signed measure, or more generally a pseudomeasure.)

This activity-type is *exemplified* on the set $R \times G$ (where G is now a subset of "real" Space-Time, $S \times T$) iff there is a measurability-preserving isometry $f: S \to S'$ and a real number c such that (8) and (9) are satisfied.

The construction of activity-types in the sense of histories is similar. With S', T' as above, an *abstract history* is a function with domain T' and range in $(R \times S') \cup \{z_o\}$ satisfying the requirements for being a history in the ordinary sense. An *activity-type* is then a measure μ' whose universe set is a measurable set of abstract histories, H'. This is *exemplified* on the set of "real" histories H iff there is an $f: S \to S'$ and a number c, as above, such that (10) and (11) are satisfied.

Finally, we may have an activity-type having both a "histories" component and a production and/or consumption component, both structures being superimposed on abstract Space-Time, $S' \times T'$. This complex activity-type is then *exemplified* on "real" sets H and $R \times G$ iff there is an f and c as above ~~which~~ that satisfy (8), (9), (10), and (11) *simultaneously*.

Scale of Activities

The question of whether there are "constant returns to scale" remains a vexed one in the economics literature. This is properly a question of technology, not description, and we therefore do not discuss it here. We do suggest, however, that much of the disagreement arises from the fact that "scale" is an ambiguous concept. In this section we shall spell out several of its possible meanings.

Given two activities, $\underline{q}$ and $\underline{q}'$, when is $\underline{q}'$ a k-fold expansion of $\underline{q}$, where $\underline{k}$ is a positive real number? All of the following answers have this in common: When $\underline{k} = 1$, they reduce to the concept defined above of $\underline{q}$ and $\underline{q}'$ being the same activity (or activity-type). Thus we are really seeking to generalize to the case where $\underline{q}$ and $\underline{q}'$ are somehow "similar" but unequal in "size".

Specifically, let $\underline{q}$ be the complex activity consisting of the measure μ over the measurable subset of histories $\underline{H}$, and λ over $\underline{R} \times \underline{G}$ ($\underline{G}$ being a measurable subset of $\underline{S} \times \underline{T}$). λ represents the production-consumption components, and may be a signed measure, a pair of measures, etc. (the following discussion is valid for all of these possibilities). Similarly, $\underline{q}'$ consists of the measure μ' over $\underline{H}'$, and λ' over $\underline{R} \times \underline{G}'$ (λ' being of the same character as λ; $\underline{G}' \subseteq \underline{S} \times \underline{T}$, and measurable).

Definition: $\underline{q}'$ is a k-fold expansion of $\underline{q}$ in the intensive sense iff there is a measurability-preserving isometry $\underline{f}: \underline{S} \to \underline{S}$, and a real number $\underline{c}$, such that

(i) (8) and (10) are valid; and

(ii) (9) is replaced by

$$\lambda'(\underline{E}) = \underline{k} \cdot \lambda\{(\underline{r},\underline{s},\underline{t}) \mid (\underline{r}, \underline{f}(\underline{s}), \underline{t} + \underline{c}) \in \underline{E}\}, \qquad (12)$$

for all measurable $\underline{E} \subseteq \underline{R} \times \underline{G}'$, and

(iii) (11) is replaced by

$$\mu'(E) = k \cdot \mu\{h | h^{c,f} \in E\}, \qquad (13)$$

for all measurable $E \subseteq H'$.

That is, the "locations" of the two activities in Space-Time have the same "size" and "shape; also the relative distribution of mass over these locations is the same, but the absolute levels on corresponding sets are k times greater for q'. If these represent two shoe factories, we would find k times as much machinery, inventories, workers, etc., crowded into the same area, turning out shoes and consuming materials at k times the rate in one of these factories as compared to the other.

This is a rather unusual conception of "scale", and we list it first only because it is the simplest of the possibilities. Indeed, one is tempted to say: "This is not a scale expansion at all: all factors, including land, must be multiplied in proportion, while the 'intensive' concept leaves the quantity of land unchanged!" We shall now try to pin down the alternative notion of scale which underlies this expostulation.

The first difficulty revolves about the concept "quantity of land". "Land" is an ambiguous term, sometimes referring to a certain class of resources which includes dirt, minerals and trees, and sometimes is a synonym for Space. Now, in the intensive scale concept, land in the first sense has been multiplied by k: the soil is k times more densely packed, etc. The protest above must therefore refer to the second meaning of the term "land". But what then is the "quantity of Space", and how does one multiply it by k?

The simplest approach is to identify "quantity of Space" with *volume* in the case of 3-space, and with *area* in two-dimensional cases, such as the plane or the surface of a sphere.[46] We shall assume that S is endowed with such a quantity measure, α, and refer to it generically as "area".

Next we need to generalize the concept of isometry.

Definition: Let (A,d) and (B,d') be two metric spaces. $f: A \to B$ is a *similarity* iff

(i) f is onto; and

(ii) there is a positive real number, m, such that

$$d'(f(x), f(y)) = m \cdot d(x,y), \text{ for all } x, y \in A.$$

m is the *dilatation* of f.

Thus f has the effect of stretching all distances by the factor m. (If $m < 1$ this is of course a shrinkage. f is an isometry iff $m = 1$.) If f is a similarity with dilatation m, then it has an inverse, $g: B \to A$, which is a similarity with dilatation $1/m$. Roughly speaking, a similarity preserves "shape" but not "size".

Let A be 3-space, and d the Euclidean metric on A; let f be a similarity from (A,d) to itself, with dilatation m. Then it may be shown that f is measurability preserving, and that

$$\mu(E) = m^3\mu\{a|f(a) \in E\}, \qquad (14)$$

for all Borel sets $E \subseteq A$, where μ is volume (that is, three-dimensional Lebesgue measure). Thus f expands volume by the cube of the dilatation. Similarly, if (A,d) is the Euclidean plane then (14), with m^2 substituted for m^3, and μ being two-dimensional Lebesgue measure, is valid. We shall confine attention to these two cases.

Again let q be the activity given by measure μ on the set of histories H and λ on $R \times G$; similarly q' is given by μ' on H' and λ' on $R \times G'$. We then define:

q' is a <u>k-fold expansion</u> of q in the <u>extensive sense</u> iff there is a real number c, and a similarity $f: S \rightarrow S$ with dilatation $k^{1/D}$ such that (8), (10), (12), and (13) are valid. (Here D is the dimensionality of Space: $D = 2$ for the plane, and 3 for 3-space).

Thus the single difference between the intensive and extensive scale concepts is that the spatial transformation has a dilatation of 1 in the former case, and of $k^{1/D}$ in the latter. The reason for this latter choice is that area (or volume, for $D = 3$) expands in exactly the same ratio as μ and λ expand on corresponding sets. Thus the <u>average density</u> of all measures with respect to the "quantity of Space" is the same for q and q' for corresponding sets. This presumably is the meaning of a "proportional expansion of all factors, including land".

If q is the normal shoe factory, and q' is a k-fold expansion of q in the extensive sense, then the workers in q' would be Brobdingnagians (if $k > 1$) or Lilliputians (if $k < 1$); all machinery and plant would expand or shrink in the same proportion. Stocks of resources and rates of output and inflow would expand by the factor k, but per unit area (or volume) would remain the same as before.

We merely mention in passing another class of "scale" concepts: those involving <u>time-dilatation</u>. In all cases discussed so far the only transformation to which Time was subjected was a simple translation: $t \rightarrow t + c$. But there could also be a scale factor: $t \rightarrow kt + c$, where k is a positive real number other than 1. The effect of this is to change the <u>speed</u> at which processes occur, the rate at which "particles" fulfil their histories. There are numerous possibilities, depending on whether Space is also subjectd to a dilatation, by the factors multiplying μ and λ, and by the relations among these four magnitudes.

An example of a time-dilatation is the relation between a film run at normal speed and the same in slow motion. (One could even have $k < 0$, which corresponds to running the film backwards).

Another concept rather different from any of the foregoing is that of scale in the <u>duplicative</u> sense. Here the expansion factor k must be a positive integer. Again let q be the activity given by μ on H and λ on $R \times G$, and q' the activity given by μ' on H' and λ' on $R \times G'$. Then

Definition: q' is a k-fold expansion of q in the duplicative sense iff there is a measurable partition of H' into k pieces — say, $H_1, \ldots, H_k$ — and of G' into k pieces — say, $G_1, \ldots, G_k$ — such that the activity q_i', defined by μ' restricted to H_i and λ' restricted to $R \times G_i$, is the same activity as q for all $i = 1, \ldots, k$, the same time-translation c serving for all $i = 1, \ldots, k$.

This definition captures the notion of the same processes running "side-by-side" — in row housing, banks of machines, or the plants of a perfectly competitive industry. The stipulation on c requires simultaneous acting out by the k-fold duplicates; this could be relaxed, to allow for staggered timing, or even for duplication by a k-fold repetition in Time.

This completes our short survey of some meanings of "scale". As mentioned, we shall return to it later with a discussion of "returns to scale" (p.).

Some Everyday Activities

In this final subsection we shall examine how various broadly defined, "one-digit" activity categories — such as mining, transportation and services — fit into the present framework. Since these categories were not designed to be so fitted, and since their definitions involve many *ad hoc* elements, we can hope at best for a broad-brush characterization, with many errors in detail.

One can classify activities from many points of view, for example, by the number of persons participating. Thus one may distinguish natural activities (no participants), private activities (one participant), and shared activities (more than one). Of the latter one may distinguish various authority structures, cooperative vs. conflictive aspects, who performs what services for whom, etc.

At the moment, however, we are mainly interested in the physical structure of activities. We shall take activities in the histories sense, and pose the problem as follows: What characterizes the defining set of histories, H, of, say, an activity classified as "construction"?

Consider transportation, for example. An ideal transportation activity is one in which all histories $h \in H$ have constant transmutation-paths, at least over the interval $[t_1, t_2]$ to which one is referring. That is, a typical "particle" may change its location in Space but not its resource-form $r \in R$. This is of course an approximation: travelers get fatigued, cargo spoils, vehicles suffer wear and tear, etc.

The foregoing approximate description applies not only to the activities customarily called "transportation" but to several others as well: utilities, such as water, gas, electricity, and sewage disposal; communications, such as telephone and broadcasting. The postal system consists mainly of transportation activities. The great bulk of everyday activities, in fact, will have a transportation sub-activity in them.

For some transportation activities it is useful to idealize even further and take $\underline{H}$ so that all its members have the same itinerary over the relevant time-interval. This makes it simply-located, and the activity may be represented by a resource-bundle (that is, a measure over a subset of $\underline{R}$) traveling over the "track" determined by the common itinerary. This approximation is good for transportation that goes in channels (roads, pipes, wires, etc.) but poor for broadcasting.

A special case of transportation is storage, the simplest of all activities. An <u>ideal storage activity</u> is one in which all histories are constants, at least over the relevant time-interval. That is, the "particles" change neither their resource-forms, $\underline{r} \in \underline{R}$, nor their locations, $\underline{s} \in \underline{S}$. This of course approximates processes in the real world which change "slowly". What is to be considered "slow" depends on one's focus of attention and scale of observation. To the historical geologist the Earth has undergone great changes, but on the human scale it has a certain massive sameness, except for changes in the weather and "minor" fluctuations such as earth-quakes and floods. Again, an economist interested in short-term business fluctuations can treat the stock of capital goods and population as constants, but this is not true for one studying economic development.

Note that when one compares ideal transportation or storage with the bundle of processes labeled "transportation" or "storage" in the real world, one must not only approximate but also abstract from certain aspects. The fuel consumed in refrigerating a warehouse does not itself satisfy the conditions for ideal storage even approximately. In simply-located transportation one focuses on the bundle being moved (including the vehicle, if any), and abstracts from the fixed plant of the transportation system: the train, but not the rails; the electric current but not the wires.

Trade, retail and wholesale, is largely a matter of transportation and storage.

Motion in Space means ~~of course~~ motion relative to the Earth, since we have conventionally taken the Earth to be fixed in Space. There is another class of activities, however, in which the essential feature resides not in motion relative to the Earth, but in the motion of the "particles" relative to each other. In particular, a *fission activity* is one in which the itineraries diverge from each other over time, and a ~~*fission*~~ fusion *activity* is one in which they converge toward each other over time.

(These characterizations are rather vague. One could distinguish further according to whether the divergence of itineraries did or did not depend on the resource-states of the histories, giving us *segregating activities* or simple *scattering activities,* respectively. Also one could go into the various ways of measuring dispersion and association of spatial distributions.[47] But this is unnecessary for the present discussion, which is impressionistic in any case.)

Going to <u>manufacturing</u>, it appears that -- very roughly, and with many exceptions -- one can classify manufacturing processes into fission activities, in which things are taken apart, or separated into component substances, and fusion activities, in which they are put together, or assembled into larger units. First-stage processing of raw materials is generally of the fission type: Crude oil is refined, ores are beneficated, crops are winnowed, logs and carcasses are chopped up. ~~(The reason is that nature presents us with things whose ingredients are mixed up in non-useful ways, and which are unwieldy in size).~~ Late stages tend to be of the fusion type: Cars are assembled, cotton is spun, woven and sewn into clothing, drugs are blended, ~~etc.~~ 48

<u>Construction</u> is a kind of fusion process ~~that is~~ distinguished by the nature of its product. This is not so much a question of size -- (supertankers and jumbo jets are larger than most buildings) -- but rather that the product is attached to the Earth; it is "real" rather than "movable" property. There are, ~~to be sure,~~ cases where it is not clear whether a given item is "real" or "movable" (e.g. fixtures, "mobile" homes), but ~~by and large~~, buildings, dams, bridges, roads, railway tracks, airports, docks, pipelines all belong to the former category, while vehicles, machines, and consumer goods belong to the latter. In summary, construction is a fusion to the Earth.

Mining is the reverse of construction. It consists of detaching pieces from the Earth. This will include not only the extraction of minerals in the conventional sense, but also the undoing of previous construction in demolition work. (Also tunneling would be considered a form of mining on this approach, just as land filling would be considered a form of construction). In summary, mining is a fission from the Earth.

What about agriculture, forestry, hunting, and fishing? These are generally classified as extractive, and indeed they have a strong mining component, as defined above. (A certain style of agriculture is known pejoratively as "soil mining"). But these increasingly tend to be run as self-sustaining processes by re-seeding, re-stocking, and fertilization, so that the "construction" aspect is becoming as important as the "mining" aspect.

This brings us to services. At first glance, this seems to be such a heterogeneous category - embracing repairs, business and personal services, professional services, entertainment and education, etc. - that no succinct property could begin to approximate it. And, indeed, this will be our contention as far as the physical structure of these processes is concerned.

Adam Smith divided workers in productive and unproductive, and it is clear from his examples of the latter - servants, lawyers, musicians, etc. - that he had in mind more or less the present day distinction between those engaged in the production of goods vs. services.[49] Although service workers have long since been admitted as contributors to the national product,[50] the tradition lingers that they produce an "intangible". This is clearly wrong in detail: Laundering is a service and there is nothing intangible about dirty laundry - or rather, about the transformation from dirty to clean laundry. What makes laundering a service is that the laundry does not own the item it is cleaning. We claim, in fact, that this characteristic, rather than any "intangibility", is what distinguishes the bulk of the activities known as services.

If true, this means that services are such not because of any physical property of the activity but because of the ownership relations among the interested parties. Hence the same activity may be either a service or not, depending on the organization of the industry. Suppose laundries operated as used car dealers do, buying dirty shirts, cleaning them, and re-selling them on the second-hand market. This may well be considered goods production. Conversely, suppose, say, copper refineries operated as follows. Miners ship their ore to the refineries without relinquishing ownership; the refined copper is then returned to the owners, who pay a fee for the service. This is completely analogous to the organization of laundries. Would not copper refining then be considered a service industry?

We shall quickly run through the major service categories, to indicate how well this characterization applies. There is no problem with repair services in general - e.g. watches, shoes, cars, radios. In all cases the owner A relinquishes possession of the item to repairman B, who fixes and returns the item to *A*.

What about *rentals* – of houses, hotel rooms, or cars, say? Let us look at the servicing relation a little more closely. Person *A* owns some items, α; *B* owns some items, β; α and β are brought together, with the result that α is benefited, in return for which *A* pays a fee to *B*. For repairs, β typically consists of the repairman himself and his tools; for laundering it consists of cleaning equipment, etc. Now the fee can be described as a rental payment for the services of β; rentals and services are two ways of looking at the same transaction. When gardener *B* trims *A*'s rosebush, one may say either that *A* rents *B*'s labor services, or that *A* relinquishes possession of his rosebush to *B*, who returns it to *A* in improved condition. In the case of house rental, β is the house itself. What is α? *A* himself and his possessions, which are provided with shelter services.

Since a person always owns his own body (in a non-slave society) any benefits to *A*'s body (including his mind) made by another person, *B*, automatically fall into the category of service activity, according to our ownership criterion. This includes the services of physicians, dentists, barbers, sex partners, and, ultimately, morticians. It includes the services of clergymen, of entertainers, and of all who provide information: Teachers, lawyers, physicians again, consultants, employment agencies, private detectives, credit bureaus, telephone answering services, and astrologers. (Perhaps this group explains the connection of services with "intangibles": one cannot see directly the changes in a person's information state or welfare level).

Most <u>government</u> activities would be services as here defined, because they provide benefits to persons and goods not owned by government.

This brief survey appears to cover ~~the great bulk of~~ most activities customarily classified as "services". Of the remainder, a number seem simply to be misclassified. (We are of course now turning the tables, and using our ownership criterion to determine what "should be" considered a service activity). From our point of view, photographers, duplicating services, and sign painters are goods producers. It is true that their products are closely tailored to individual clients, but the same is true of much housebuilding, job shop work, printing, and other activities classified as goods production. Similarly, a lawyer writing up a will or a contract is engaged in goods production. The most important misclassified industry is advertising, whose product -- (again tailored to individual clients) -- is "advertising copy", consisting of jingles, skits, blurb, etc.

Services will be discussed further in connection with rental markets in chapter 6.

2.8. Multi-layer Measures

For a general measure space (A,Σ,μ), nothing specific needs to be assumed about the nature of the points of A. In our applications, A has variously been a product space built up from R, S, and T, or a space of functions (histories) whose domain and ranges are built up from these, or subsets of the foregoing, etc. We now ~~briefly~~ discuss some cases in which the points of universe set A are *themselves* measures over some other measurable space.

Let us spell this out. Given a fixed measurable space, (B,Σ), let $\mathbb{M}$ be the set of all measures over it. ($\mathbb{M}$ could also be the set of all signed measures or all pseudomeasures; the discussion would be unaffected). Now consider a measure space $(\mathbb{M},\Sigma',\mu)$, in which $\mathbb{M}$ itself plays the role of universe set. We shall refer to this as a *two-layer measure*. Next, suppose B itself is a set of measures over still another space, so that each member of $\mathbb{M}$ is itself a two-layer measure; then $(\mathbb{M},\Sigma',\mu)$ will be referred to as a *three-layer measure*. This clearly extends to any finite $n = 1, 2, \ldots$.

We consider some ways in which such multi-layer structures arise in applications. Let us first bring in the factor of *uncertainty*. We have noted that, in principle, the measure space of histories, (Ω,Σ,μ), provides a complete description of the world for social science purposes. But ~~of course~~ one never knows exactly what the measure μ is. It is desirable, then, to try to represent states of relative ignorance or degrees of belief concerning the true measure, μ.

We shall adopt a "Bayesian" point of view, according to which "state of belief" is representable as probability measure over the universe set of "possible worlds".[51] Specifically, a state of belief is given by $(\mathcal{M}, \Sigma', \pi)$, where $\mathcal{M}$ is the set of all measures over the measurable space of histories, (Ω, Σ), Σ' is a sigma-field on $\mathcal{M}$, and π is a probability measure with domain Σ'. For any $\underline{E} \in \Sigma'$, $\pi(\underline{E})$ is the probability (= "degree of belief") that the true mass distribution, μ, over the space of histories, belongs to the set of measures $\underline{E}$. This is a two-layer measure.

How is Σ' determined? The heuristic principle we have used before states that all sets which are "conceptually observable" should be considered measurable. Here the equivalent principle would seem to be: Any set of measures which is "sufficiently simple" so that a mind could, conceptually, hold a degree of belief concerning it should be considered measurable. This is rather vague, and is best explained by examples. If $\underline{F}$ is a measurable set of histories, and $\underline{c}$ a number, the event: "the total mass concentrated on the histories in $\underline{F}$ exceeds $\underline{c}$" would appear to be one to which a degree of belief could be attached. This means that the set of measures

$$\{\mu \mid \mu(\underline{F}) > \underline{c}\} \qquad (18)$$

is to be considered measurable (that is, belongs to Σ') for all $\underline{F} \in \Sigma$, all real $\underline{c}$. In particular, suppose $\underline{F}$ is the set of histories originating in subset G of R × S × T; then the the probability, π, attached to the set (1) gives the degree of belief that total births or production in G exceeds the value c.

This example provides an illustration of how the probability π can be induced onto simpler spaces. Let F retain its meaning of the set of histories originating in fixed set G, and consider the function with domain M that assigns μ(F) to measure μ. This induces a probability measure on the real line, which is exactly the state of belief concerning production in set G. This induction process is completely analogous to the many examples in section 2.5 of the induction of μ on the space of histories (Ω,Σ) onto simpler spaces.

The case of perfect certainty, with a known measure μ_o over (Ω,Σ), may be identified with the special case of the probability measure (M,Σ',π) in which π is simply concentrated with all mass at the "point" $\mu_o \in$ M.

For a second example, consider the structure of the Resources set R. Taking people types as points of R, a complete specification of a person r ∈ R will include his mental state, in particular his state of knowledge. Let us assume for the moment that r describes a person in a state of complete certainty. His state of knowledge will then include a description of the world, which is represented as a measure over the space of histories, Ω.

This lends again to something resembling a two-layer measure, for some of the points of R have an internal structure involving measures, while the overall descriptive measure is on a universe set built up in part from R.

In fact, once one admits into R structures which involve measures over universe sets involving R, one appears to be led to "infinite-layered" structures. The reason is that a person's state of knowledge will itself be (at least) two-layered, since it involves knowledge of other states of knowledge; and one cannot stop at any finite number of layers.

Whether one can build a useful (or even consistent) theory from such an infinite regress remains to be seen. There is one consideration which simplifies things, however. The measure space (Ω,Σ,μ) gives a complete description of the world. But any person, even one in a state of perfect certainty, will have a limited capacity to assimilate information. This limitation may be represented formally by replacing Σ by a small sub-sigma-field $\Sigma' \subseteq \Sigma$, yielding an aggregation of the original measure and losing detail. Knowledge of other people's states of knowledge (and of one's own past and future states) would be in terms of an even smaller sigma-field $\Sigma'' \subseteq \Sigma'$, hence even more aggregative and sketchy, etc.

Finally, mental states of uncertainty can be represented, as above, by a probability measure over a universe set M of physical measures. We again get an infinite regress, in the form: A's degree of belief concerning B's degree of belief. ....[52]

In this book we shall not have occasion to use multi-layer measures in any of the interpretations just discussed. We shall, however, use them in another way, as a representation of technology. Here M will be the set of "basic feasible activities", and feasible activities in general will be measures over M. For detailed discussion see chapter 4, section 5.

FOOTNOTES

[1] Certain new mathematical results will be given later. Most of the mathematical material in this chapter is standard, except possibly for terminology.

---

[2] The following books are recommended for readers wishing to go beyond the necessarily sketchy outline of measure theory presented in this chapter. They are roughly in decreasing order of difficulty:

N. Dunford and J. T. Schwartz, Linear Operators, vol. I, chapter III (Wiley-Interscience, New York, 1958);

S. Saks, Theory of the Integral (Stechert-Hafner, New York, 1937);

H. Hahn and A. Rosenthal, Set Functions (University of New Mexico Press, Albuquerque, 1948);

P. R. Halmos, Measure Theory (Van Nostrand, Princeton, 1950);

S. K. Berberian, Measure and Integration (Chelsea, New York, 1970);

M. E. Munroe, Measure and Integration (Addison-Wesley, Reading, Mass., 2nd edition, 1970);

A. E. Taylor, General Theory of Functions and Integration (Blaisdell, New York, 1965), (second half of book).

In addition, books on the theory of functions of a real variable will generally have pertinent material. But the books above have been deliberately chosen for their abstract approach to the subject, which is ~~the approach~~ suitable for the applications we wish to make.

~~It should be noted that~~ terminology has not been fully standardized, so that these books differ among themselves, and with ~~the present~~ book.

---

[3] ~~In fact,~~ any dense subset of the reals could be used.

---

[4] The numbers $\pm\infty$ in the extended real number system should not be confused either with "indeterminate forms" in calculus, which are just abbreviations for certain limit operations, or with infinite cardinal numbers in set theory.

---

[5] It also often involves a <u>distortion</u>, to squeeze the facts into the categories of the formal system — e.g. the assumption of perfect vacua, ideal gases, pure substance, and, in social science, of perfect competition, "lightning calculation", "ideal types", rationality, economic man, political man, libidinal man, etc.

---

[6] Common borderlines may be thought of as "no-man's land" which belongs to none of the abutting countries.

---

[7] These correspond roughly to Norman Campbell's "fundamental ~~magnitudes~~". cf. N. R. Campbell, <u>Foundations of Science</u> (Dover, New York, 1957), ~~chapter~~ X.

---

58 [8] It could be more inclusive than the Borel field. But the attempt to extend Lebesgue measure to the class of <u>all</u> subsets of the real line runs into counter-intuitive paradoxes [at least in the realm of "standard" measure theory, the kind used in this book and everywhere else until very recently. But cf. A. R. Bernstein and F. Wattenberg, "Nonstandard Measure Theory", pp. 171–185 in *Applications of Model Theory to Algebra, Analysis and Probability*, W. A. J. Luxemburg, ed. (Holt, Rinehart and Winston, New York, 1969).]

---

59 [9] In fact, the great contribution of Lebesgue consists in sensing and systematically developing the consequences of countable additivity, as opposed to the earlier finitely-additive "Jordan content".

---

59 10. [9a] De Finetti has argued ~~strongly~~ that subjective probabilities should be only finitely additive. But the measures we are discussing represent physical magnitudes, not degrees of belief (with the exception of section 8), and his strictures do not apply to them. B. de Finetti, *Probability, Induction and Statistics* (Wiley, New York, 1972).

60 [10] ~~Actually~~ this reduces to the countable case as follows. If the set $I' \subset I$ on which f is positive is uncountable, ~~then~~ the summation of f equals ∞; if I' is countable, ~~then~~ the summation of f is the same as that of f restricted to I'. But this reduction does not detract from the intuitive appeal of a single definition covering all cases.

[11] How do we know that the territories occupied by countries are in fact Borel sets (that is, members of the Borel field)? An element of convention enters here. Without attempting a rigorous discussion, it may be said that any real-world region presenting itself as an observational unit is empirically indistinguishable from some Borel set. Thus taking it to be a Borel set is a mathematical convenience which does no violence to the facts.

[12] It is clear from this list, by the way, that the term "resource" is misleading. Other possible terms, such as "substance", "essence", "quiddity", or "quality", seem even worse. One should keep in mind, then, that "resource" is a general neutral term embracing people-types as well as goods-types, and "illth" as well as "wealth".

[13] Spatial "configurations", as we shall see, can be represented by measures.

[14] No confusion should arise between $\sum$ as used here to indicate summation, and boldface Σ to represent a sigma-field.

[15] Note that $\Pi\Sigma_i$ and $\Sigma_1 \times \ldots \times \Sigma_n$ are not the cartesian products of the family $(\Sigma_i)_{i \in I}$. No confusion should result from this ambiguous notation.

[16]There is ~~obviously~~ a strong element of convention in this statement. One can hardly identify a unique moment at which a person pops into or out of existence (e.g., ~~one~~ we might start with conception rather than birth; ~~one has~~ there are problems with resuscitation ~~and~~ suspended animation, etc.) The inclusion of the endpoints t_1, t_2 in the interval is also obviously a pure convention.

---

[17]This distinction hinges on the scale of observation. All ~~of~~ the "continuous" resources mentioned reveal a "granular" structure under the microscope. Conversely, from a large-scale point of view it may be useful to think of people, say, as being continuously distributed, as when one speaks of "population density" or "migration flow".

---

[18]For people, the "number of entities" approach is almost universal. No political system is organized on the principle, "one pound, one vote"!

---

[19]This point will be elaborated on after we have defined the concept of "integral."

---

[20]~~There is~~ an element of convention is involved in defining E. ~~For example~~ e.g., is E to include the entire volume enclosed by the building, or just the shell? We suppose this has been decided.

---

21 In particular, the conpept of "congruent measures"; ~~see p. .~~

22 We have deliberately refrained from introducing the metrical or topological concepts that would be needed for this. These notions play a decidedly secondary role in this book, and we have therefore concentrated on building up the theory of measure *per se*, which does not depend on them.

23 ~~There is no convenient way to introduce continuity at the moment of birth or death.~~ A history is said to be continuous iff it is continuous at all instants of time except for ~~these two~~, and continuous from the (fuGure, past) at (birth, death), respectively.

[24] Can one go further, and decompose the rural population into a "two-dimensional" part (e.g. the farm population), and a "one-dimensional" part (e.g. population living along roads, rails, or rivers), and perhaps a residual? The answer is yes, with the aid of the Lebesgue decomposition theorem, which we shall not cover in this book. For this "dimensional decomposition" see H. Hahn and A. Rosenthal, Set Functions, 106-109.

---

[25] Later on we will allow both μ and f to take on negative values. Unless otherwise noted, all functions from now on will have their range in the extended real numbers. (This, of course, does not mean they must take on infinite values, only that they may do so.)

---

[26] There are a large number of seemingly different definitions of the integral in the literature. Most of these are either equivalent to (4) or minor variants of it.

---

[27] After W. H. Young, 1905. The general integral (4) is essentially due to M. Fréchet, 1915.

---

[28] For readers troubled by this cavalier addition of heterogeneous units - often said to be "invalid" - it should be mentioned that it is clearer to think of the measurement unit as being part of the definition of the concept, the measurement number itself being "pure". Thus, "the length of this bar in

meters: 3.7", rather than "the length of this bar: 3.7 meters". cf. R. Carnap, Introduction to Symbolic Logic and Its Applications, W. H. Meyer and J. Wilkinson, trans. (Dover, New York, 1958), p. 169. In any case, the present treatment shows how to work with heterogeneous units with safety and convenience.

---

[29]We shall discuss higher dimensional Lebesgue measure below.

---

[30]By the uniqueness theorem, any function which is identical to p_1/p_0 except for a set of μ_0-measure zero is also an Radon-Nikodym derivative. In this case, as in many others, the derivative p_1/p_0 is more "natural" than the other functions μ_0-equivalent to it.

---

[31]Densities correspond roughly to Campbell's "derived magnitudes". See N. R. Campbell, Foundations of Science, Chapter X, and footnote 7 above.

---

[32]In rate-of-return calculations the density is unknown, but its derivation is not comparable to our procedure above. Incidentally, (18) illustrates the alternative notation for the integral given in (2): When several letters are floating around, it clarifies which measure and integrand one is referring to.

---

[33] P. R. Halmos, Measure Theory, page 22. The proof of the following theorem may be found in Halmos, Chapter III, or in J. von Neumann, Functional Operators, vol. I, Measures and Integrals (Annals of Mathematics Studies, #21, Princeton University Press, Princeton, N.J., 1950).

[34] In the literature, conditional measures arise mainly in probability theory, e.g., in J. L. Doob, Stochastic Processes (Wiley, New York, 1953), Appendix. Note that "conditional probability" is often used in a quite different sense than the one employed here.

[35] On distribution functions see J. von Neumann, Functional Operators, I, Measures and Integrals, (Annals of Mathematics Studies #21, Princeton University Press, Princeton, N.J., 1950), pp. 160-172; H. Cramér, Mathematical Methods of Statistics (Princeton University Press, Princeton, N.J., 1946), pp. 77-82. Definitions vary from one author to another.

[36] We shall see later that the method for solving these problems is closely related to work of Ramsey, Weizsäcker, and others on the evaluation of infinite development programs. In fact, it incorporates these "overtaking criteria" as special cases. See Sections 3.3 and 3.4.

[37] $\mu^+(\underline{E})$ must include exports from $\underline{E}$ to itself, not merely to $\underline{A}\backslash\underline{E}$, for otherwise μ^+ would not be an additive set-function; similarly for $\mu^-(\underline{E})$. These sketchy statements will be elaborated when we come to discuss transportation and transhipment, chapter 7.

---

[38] When generalizing even further, to pseudomeasures, one does not have to do this. See chapter 3.

---

[39] One of the advantages of pseudomeasures is that the proviso concerning expressions of the form $\infty - \infty$ may be dropped. When (47) is suitably generalized, the (indefinite) integral of <u>any</u> measurable real-valued function with respect to <u>any</u> pseudomeasure is well-defined. See chapter 3.

---

[40] This is the first occasion on which non-measure-theoretic -- specifically, metric -- concepts are being used.

---

[41] One typically postulates certain further relations between $\underline{d}$ and Σ. However, this is not necessary for the present discussion.

---

[42] There is one fine point that at least deserves footnote mention. The distributions representing the left- and right-hand gloves of a pair are measure-congruent, but cannot be transformed into one another by a rigid motion: one must be

turned "inside-out". We could therefore insist that sameness of type requires not only measure-congruence, but preservation of "parity" or "orientation". Since this condition does not seem important for social science problems, we pass over it without further discussion.

---

[43]The "same" melody in two different keys, or the replacement of pine by spruce in a house, are examples of shifts among resource types analogous to interregional shifts. But these apply only within small "homogeneous" subsets of R.

---

[44]We have already mentioned some problems concerning which subsets are to be considered measurable in the set of histories. Here we merely assume implicitly that all sets mentioned are in fact measurable.

---

[45]Also called a Weberian activity, after Alfred Weber. See 9.4.

---

[46]How to define these measures in the case of more complicated manifolds is itself a rather difficult problem into which we shall not delve. See L. Cesari, Surface Area (Annals of Mathematics Studies, #35, Princeton University Press, Princeton, N.J., 1956), or T. Radó, Length and Area (American Mathematical Society Colloquium Publications, vol. 30, Providence, R.I., 1948).

---

[47] D. S. Neft, *Statistical Analysis for Areal Distributions* (Monograph Series #2, Regional Science Research Institute, Philadelphia, 1966).

---

[48] Fission and fusion activities correspond roughly to Beverly Duncan's "processing" and "fabricating" industries, and even more roughly to Alfred Weber's "material-oriented" and "market-oriented" industries. See. O. D. Duncan, W. R. Scott, S. Lieberson, B. Duncan and H. H. Winsborough, *Metropolis and Region* (John Hopkins Press, Baltimore, 1960), 57-58 and Chapter 7, and E. M. Hoover, *The Location of Economic Activity* (McGraw-Hill, New York, 1948), 31-38, respectively.

---

[49] *Wealth of Nations*, Book II, Chapter III.

---

[50] In the West, not in the Communist world.

---

[51] See H. E. Kyburg, Jr., and H. E. Smokler, editors, *Studies in Subjective Probability* (Wiley, New York, 1964), especially the essays by B. de Finetti and B. O. Koopman.

---

[52] Even an omniscient Deity would have need for probability concepts, to represent the states of mind of the less-than-omniscient creatures inhabiting the world.
