formally weak notions and because the analyses of such concepts seem best left in other hands [16].

References

1. C. Shannon and W. Weaver, *Mathematical Theory of Communication*, pp. 9–18. Urbana: Univ. of Illinois Press, 1949.
2. N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, pp. 1–13. New York: Wiley, 1949.
3. H. Feigl, De Principiis Non Disputandum ... ? in *Philosophical Analysis* (M. Black, ed.), pp. 113–149. Englewood Cliffs: Prentice-Hall, 1963.
4. R. Carnap, *Logical Foundations of Probability*, 2nd ed. Chicago: Univ. of Chicago Press, 1962.
5. C. Lewis, *An Analysis of Knowledge and Valuation*. LaSalle: Open Court, 1946.
6. M. Black, *Margins of Precision*, pp. 57–208. Ithaca: Cornell Univ. Press, 1970.
7. I. J. Good, *Probability and the Weighing of Evidence*. London: Griffin, 1950.
8. W. Kneale, *Probability and Induction*. London and New York: Oxford Univ. Press (Clarendon), 1949.
9. H. E. Kyburg, Jr., *Probability and Inductive Logic*. New York: Macmillan, 1970.
10. J. R. Lucas, *The Concept of Probability*. London and New York: Oxford Univ. Press (Clarendon), 1970.
11. W. C. Salmon, *The Foundations of Scientific Inference*. Pittsburgh: Univ. of Pittsburgh Press, 1966.
12. L. J. Savage, *Foundations of Statistics*. New York: Wiley, 1954.
13. G. H. von Wright, *The Logical Problem of Induction*, 2nd rev. ed. New York: Barnes & Noble, 1957.
14. A. P. Dempster, On the Difficulties Inherent in Fisher's Fiducial Argument, *J. Amer. Stat. Assoc.* 59, 56–66, 1964.
15. R. B. Braithwaite, *Scientific Explanation*, pp. 151–191, reprint. New York: Harper & Brothers, 1960.
16. S. Toulmin, *The Uses of Argument*. London and New York: Cambridge Univ. Press, 1958.

II

Axiomatic Comparative Probability

IIA. Introduction

The concept of comparative probability (CP) exemplified by the statement "A is at least as probable as B" has thus far received very little attention from either engineers, scientists, probabilists, or philosophers. Grounds for a greater interest in this neglected concept include the following points.

(1) CP provides a more realistic model of random phenomena when we have too little prior information and data to estimate quantitative probability reasonably.

(2) CP provides a wider class of models of random phenomena than does the usual quantitative theory.

(3) CP illuminates the structure of quantitative probability, and especially the Kolmogorov axioms, by providing a base from which to derive quantitative probability.

(4) CP appears to be a sufficiently rich concept to support a variety of significant applications.

With respect to (1), the logical relation between CP and quantitative probability is such that we are always more confident about estimates of the former than of the latter; for example having observed that 10 tosses of a strange coin resulted in 7 heads, we are more justified in asserting that "heads are more probable than tails" than in asserting "the probability of heads is .7." Point (2) refers to the curious phenomenon that there exist relatively simple examples of what we consider to be valid CP statements that are incompatible with any representation in the usual quantitative theory. One characteristic of these new models of random phenomena is that they do not admit of indefinitely repeated independent performances. Point (3) will, we hope, be vindicated by the discussion to follow here and in Chapter III. Point (4) is open to challenge. There has been very little work attempting to apply CP. While an application to decision making is given in Section IIG, clearly much remains to be done.

Our development of CP derives from attempts to answer the following questions.

(1) What is the formal, axiomatic structure of the relation "at least as probable as"?
(2) What is the relation between CP and quantitative probability as formalized by Kolmogorov?
(3) What are definitions of associated concepts such as comparative conditional probability, comparative independence, and expectation?
(4) How can we estimate or measure comparative probability so as to generate a useful empirical theory of probability?
(5) How can we use CP in decision-making and inference?

Questions (1) and (2) are examined in Sections IIB, IIC, and IID. We treat (3) in Sections IIE, IIF, and IIH and (5) in Section IIG. Remarks on the interpretive basis (4) are deferred to Section IVK for a relative-frequency basis, Section VI for a computational complexity basis, and Section VIIC for a logical basis. Before we can expect to resolve the measurement problem of (4) we will have to advance the state-of-the-art in decision making referred to in (5).

The results we present in this chapter are incomplete in many respects. Much more research will be required before comparative probability can take what we judge to be its highly significant rightful place in the analysis of random phenomena.

IIB. Structure of Comparative Probability

1. Fundamental Axioms That Suffice for the Finite Case

Comparative probability, hereafter abbreviated as CP, is a binary relation between events, or their representations as sets or propositions, that we denote by "$A \gtrsim B$" and read as "event A is at least as probable

as event B"; alternatively, we may write "$B \lesssim A$" which can be read as "event B is no more probable than event A." We define equally probable events, "$A \approx B$" to be read "event A is as probable as event B," by $A \approx B$ if $A \gtrsim B$ and $B \gtrsim A$. Finally, we define "$A \succ B$" to be read "event B is not as probable as event A" by $A \succ B$ if $A \gtrsim B$ and false $B \gtrsim A$. In the sequel the events are represented as elements of a field $\mathscr{F}$ of subsets of a set Ω (see Section IIIA) and are denoted by capital roman letters. The following five axioms partially characterize the family of CP relations $\gtrsim$ on $\mathscr{F}$.

C0. (*Nontriviality*) $\Omega \succ \varnothing$, where $\varnothing$ is the null or empty set.

C1. (*Comparability*) $A \gtrsim B$ or $B \gtrsim A$.

C2. (*Transitivity*) $A \gtrsim B, B \gtrsim C \Rightarrow A \gtrsim C$.

C3. (*Improbability of impossibility*) $A \gtrsim \varnothing$.

C4. (*Disjoint unions*) $A \cap (B \cup C) = \varnothing \Rightarrow (B \gtrsim C \Leftrightarrow A \cup B \gtrsim A \cup C)$.

Axioms C1 and C2 establish that the CP relation $\gtrsim$ is a linear, complete, or simple order. The requirement that all events be comparable is not insignificant and has been denied by many careful students of probability including Keynes and Koopman (see Section VIIC). Although we will not explore the matter, it is possible that the concept of CP should allow for an indifference (symmetric, reflexive, intransitive) relation between events. Axioms C3 and C4 would presumably be widely assented to as essential to the characterization of "at least as probable as." These five axioms appear to suffice for the cases where $\mathscr{F}$ is a finite set. Yet as we shall see in Section IIC they admit CP orders that need not be compatible with the usual probability theory.

The consequences of these five axioms include the following easily verified reasonable properties of a CP relation.

(a) $\Omega \gtrsim A$.

(b) $A \gtrsim B \Rightarrow \bar{B} \gtrsim \bar{A}$.

(c) $A \supseteq B \Rightarrow A \gtrsim B$.

(d) $A \gtrsim B, \quad C \gtrsim D, \quad A \cap C = \varnothing \Rightarrow A \cup C \gtrsim B \cup D$.

(e) $A \cup B \gtrsim C \cup D, \quad C \cap D = \varnothing \Rightarrow$ either $A \gtrsim C$, $A \gtrsim D$, $B \gtrsim C$, or $B \gtrsim D$.

If $\mathscr{F}$ is required to be merely a A-field (see Section IIIB), then the preceding consequences of C0–C4 need not hold. In this case (d) might be an appropriate replacement for C4 [1].

2. Compatibility with Quantitative Probability

A function P is said to agree with, or represent, an order relation $\gtrsim$ if

$$P : \mathscr{F} \Rightarrow R^1, \qquad A \gtrsim B \Leftrightarrow P(A) \geqslant P(B).$$

When $\gtrsim$ is a CP relation, then we will refer to P as a quantitative probability. Axioms C0–C4 do not permit us to conclude that for any CP relation $\gtrsim$ satisfying C0–C4 there exists some quantitative probability.

Counterexample (Lexicographic order) [2]. Let $\Omega = [0, 1]$ be the Borel field of subsets of Ω, $l(A)$ the Lebesgue measure (length) of A, μ a measure dominated by l which for the sake of illustration we take to have the triangular density 2ω over $[0, 1]$. We now define a relation $\gtrsim$ by: $A \gtrsim B$ if either $l(A) > l(B)$ or $l(A) = l(B)$ and $\mu(A) \geqslant \mu(B)$. It is easily seen that $\gtrsim$ satisfies C0–C4. To verify that there is no agreeing P we consider the sets $A_x = (1 - x, 1)$ and $B = (0, x)$ where $0 < x < 1$ for any x. It is evident that $l(A_x) = l(B_x) = x$, $\mu(A_x) = x(2 - x)$, $\mu(B_x) = x^2$, and for $x < x'$ we have $B_x \prec A_x \prec B_x' \prec A_x'$. Hence, for any agreeing P we must have $(P(B_x), P(A_x))$ is a nonnull interval and $x \neq x'$ implies that $(P(B_x'), P(A_x')) \cap (P(B_x), P(A_x)) = \varnothing$. Thus we have established a one-to-one correspondence between real numbers x in $(0, 1)$ and disjoint intervals in R. However, this is a contradiction, for there are uncountably many x and countably many disjoint intervals. Therefore, there does not exist P agreeing with the lexicographic order $\geqslant$. ∎

In order to propose a fifth axiom to guarantee the existence of a quantitative probability we must first recall a few terms from topology [3]. A topological space $(\mathscr{F}, \mathscr{T})$ is a set $\mathscr{F}$ and a collection $\mathscr{T}$ of open subsets of $\mathscr{F}$; $\mathscr{T}$ is the topology. A base $\mathscr{B}$ for $\mathscr{T}$ is a collection of subsets of $\mathscr{F}$, $\mathscr{B} = \{B_\alpha\}$, such that $T \in \mathscr{T}$ is a union of sets in the base. A subbase for $\mathscr{T}$ is a collection of subsets of $\mathscr{F}$ such that finite intersections between elements of the subbase generate the base. A countable base is a base with a countable number of elements. The order topology corresponding to the complete, transitive, irreflexive relation $\prec$ is the topology with subbase having as elements $\{A : A \prec B\}$, $\{A : B \prec A\}$ for any $B \in \mathscr{F}$. Finally, a set $\mathscr{D}$ is order dense if

$$(\forall A, B \in \mathscr{F} - \mathscr{D})\, A \prec B \Rightarrow (\exists D \in \mathscr{D})\, A \prec D \prec B.$$

We may now state the following equivalent axioms.

C5a. $(\mathscr{F}, \mathscr{T})$ has a countable base.

C5b. $(\mathscr{F}, \mathscr{T})$ has a countable order dense set.

The justification for these axioms is contained in the following theorem.

Theorem 1. If $\gtrsim$ satisfies C1–C4, then it admits of a quantitative representation P if and only if $\gtrsim$ also satisfies C5, where C5 can be taken to be either C5a or C5b.

Proof. All proofs of results in this chapter can be found in the Appendix to this chapter. ∎

Theorem 1 may be of value for other axiomatizations of $\gtrsim$ when there are infinitely many equivalence classes of events ($\mathscr{F}/\approx$ is infinite); C5 is trivially satisfied when $\mathscr{F}/\approx$ is finite. A possibly more intuitive sufficient condition for the existence of agreeing P, involving an axiom of monotone continuity for $\gtrsim$, is stated in Subsection IIB4.

Axioms C1–C5 guarantee the existence of some quantitative probability P agreeing with the comparative probability relation $\gtrsim$. This P is not unique. Any strictly increasing function f of P, $P' = f(P)$, yields a quantitative probability agreeing with $\gtrsim$. Hence we can, if we wish, choose P such that $P(\Omega) = 1$ and $P(F) \geqslant 0$, thereby satisfying the first two Kolmogorov axioms. However, we have said nothing about the concept of comparative probability that would necessitate these choices.

3. An Archimedean Axiom

Modifying a suggestion of Luce [4], we will propose and accept an Archimedean axiom whose purpose is to rule out those $\gtrsim$ which admit infinitely, or infinitesimally, probable events. We first introduce a definition.

Definition. $\{A_i\}$ is an upper scale for A if $A_0 = \varnothing$, $(\forall j \geqslant 0)(\exists B_j, C_j)$ $C_j \gtrsim A$, $B_j \gtrsim A_j$, $C_j \cap B_j = \varnothing$, $A_{j+1} \gtrsim B_j \cup C_j$.

In essence, an upper scale for A is any sequence of sets in which the "gap" between successive terms is of "size" at least A. It is immediate that any subsequence of an upper scale is again an upper scale, and that if $A \succ B$, then for any upper scale for A there is one at least as long (as many terms) for B. If $\gtrsim$ in the definition of an upper scale is replaced by $\approx$, then we call the resulting scale a standard series; the gaps between successive terms in a standard series are exactly of size A. We can now propose

C6. If $A \succ \varnothing$, then all upper scales for A are of length finite.

Our version of the Archimedean axiom is strictly stronger than the one proposed by Luce employing standard series in place of upper scales. Consider the following example:

$\Omega = \{\omega_i\}$, $A \in \mathscr{F}$ if either A or $\bar{A}$ is a finite subset of Ω,

$$(\forall n)\ \omega_n \succ \bigcup_{j<n} \omega_j .$$

It is easily verified that the above-defined $\gtrsim$ satisfies C0–C5 but not C6. For C5 we need only note that $\mathscr{F}$ is a countable collection. For the failure of C6 we observe that $\{A_n = \bigcup_{i \leqslant n} \omega_i\}$ is an infinite-length upper scale for ω_1. However, due to an absence of nontrivial equivalences, all standard series for ω_1 are of length 1.

The independence between C0–C5 and C6 is established in

Theorem 2.

(a) C0–C5 $\not\Rightarrow$ C6,
(b) C0–C4, C6 $\not\Rightarrow$ C5.

Upper scales and standard series can have some unintuitive properties. While at first glance it appears reasonable to think of A_n in a standard series for A as n times as probable as A, there can exist standard series $\{A_i\}$ and $\{B_i\}$ for A such that

$$(\exists j)\ A_j \approx B_j , \qquad A_{j+1} \approx B_{j+1} , \qquad (\exists i) \text{ false } A_i \approx B_i .$$

4. An Axiom of Monotone Continuity

This axiom was suggested by Villegas [5] in connection with the existence of countably additive agreeing P. For convenience we denote by $A_i \downarrow A$ the property of a countable set $\{A_i\}$ that

$$(\forall i)\ A_i \supseteq A_{i+1} , \qquad \bigcap_{i=1}^{\infty} A_i = A.$$

C7. (*Monotone continuity*) $(\forall i)\ A_i \gtrsim B,\ A_i \downarrow A \Rightarrow A \gtrsim B.$

Unlike C0–C6, we are inclined to view C7 as attractive but not necessary for a characterization of $\gtrsim$. The utility of C7 is apparent from

Theorem 3.

(a) C0–C6 $\not\Rightarrow$ C7,
(b) C0–C4, C7 $\Rightarrow$ C5 and C6.

5. Compatibility of CP Relations

Unlike the situation in the usual quantitative probability theories, it is not the case that every pair of CP experiments can be considered as arising from a single underlying CP experiment. To elucidate this we introduce a

Definition. $\mathscr{E}_1 = (\Omega_1, \mathscr{F}_1, \gtrsim_1)$ and $\mathscr{E}_2 = (\Omega_2, \mathscr{F}_2, \gtrsim_2)$ are strongly compatible if there exists $\mathscr{E} = (\Omega, \mathscr{F}, \gtrsim)$ and random variables x_1, x_2,

$$x_i : \Omega \to \Omega_i , \qquad (\forall A \in \mathscr{F}_i)\ x_i^{-1}(A) \in \mathscr{F},$$

such that x_i induces $\mathscr{E}_i$; that is,

$$x_i(\Omega) = \Omega_i , \qquad A \gtrsim_i B \Leftrightarrow x_i^{-1}(A) \gtrsim x_i^{-1}(B).$$

As is discussed in Section IIC, there are CP relations that do not admit of any additive agreeing P. It is easily shown under C0–C4 and C7 that no such CP relation is strongly compatible with any CP relation for which there is a countably additive agreeing P whose range contains an interval. A weaker version of compatibility is given in the following.

Definition. $\{\mathscr{E}_\alpha\}$ are mutually weakly compatible if there exist $\mathscr{E}$ and $\{x_\alpha\}$ satisfying: x_α has domain $D_\alpha \in \mathscr{F}$, $x_\alpha(D_\alpha) = \Omega_\alpha$, $\bigcap_\alpha D_\alpha \succ \varnothing$,

$$(\forall A \in \mathscr{F}_\alpha)\ x_\alpha^{-1}(A) \in \mathscr{F}, \qquad A \gtrsim_\alpha B \Leftrightarrow x_\alpha^{-1}(A) \gtrsim x_\alpha^{-1}(B).$$

The requirement of weak compatibility allows for the possibility that some outcomes of, say, $\mathscr{E}_1$ preclude the performance of $\mathscr{E}_2$. This is a very realistic possibility, although one which the quantitative theory had no need to consider. The question of compatible random experiments is a novel one and deserves closer examination than we have given it.

IIC. Compatibility with Finite Additivity

1. Introduction

With the structure thus far exposed for CP we can assert

Theorem 4. If $\gtrsim$ satisfies C0–C5, then there exists P agreeing with $\gtrsim$, and there exists a function G of two variables such that

$$A \cap B = \varnothing \Rightarrow P(A \cup B) = G(P(A), P(B)).$$

Furthermore, $G(x, y)$ is symmetric $[G(x, y) = G(y, x)]$, strictly increasing in x, $G(x, P(\phi)) = x$, and associative $[G(G(x, y), z) = G(x, G(y, z))]$.

Acquaintance with the theory of associative functional equations [6] might lead one to expect that by a suitable increasing transformation of P to P' we might find a corresponding G' such that $G'(x, y) = x + y$. That this is not always possible is evident from a simple counterexample due to Kraft *et al.* [7].

Counterexample. Let $\Omega = \{a, b, c, d, e\}$ and $\mathscr{F}$ contain all subsets. Consider the ordering

$$\begin{aligned} &\varnothing \prec a \prec b \prec c \prec ab \prec ac \prec d \prec ad \prec bc \prec e \prec abc \prec bd \prec cd \prec ae \prec abd \\ &\prec be \prec acd \prec ce \prec bcd \prec abe \prec ace \prec de \prec abcd \prec ade \prec bce \prec abce \\ &\prec bde \prec cde \prec abde \prec acde \prec bcde \prec abcde = \Omega. \qquad (*) \end{aligned}$$

The relation $\prec$ defined above clearly satisfies C0–C5 and hence admits of a quantitative P. However, there is no such P that is finitely additive. To verify this, assume to the contrary that there exists an additive P assigning probabilities $P(a) = A$, $P(b) = B$, $P(c) = C$, $P(d) = D$, $P(e) = E$. We then see from $(*)$ that

$$A + C < D; \qquad A + D < B + C; \qquad C + D < A + E.$$

Adding these three inequalities together and subtracting repeated terms yields

$$A + C + D < B + E.$$

Hence $acd \prec be$. However, this is contradicted by $(*)$. Thus there cannot exist an additive quantitative probability agreeing with $\prec$ defined by $(*)$. ∎

2. Necessary and Sufficient Conditions for Compatibility

Necessary and sufficient conditions that ensure that $\gtrsim$ admits of a not necessarily unique, finitely additive representation when Ω is finite have been given first by Kraft *et al.* and then by Scott [8]. Scott's treatment seems preferable and can be more easily extended to infinite Ω. After presenting Scott's theorem we will argue that the hypothesis of the theorem is not an acceptable axiom for CP. It appears that CP relations are not reasonably restricted to only those compatible with additive probability. Nevertheless, it is desirable to have intuitively appealing sufficient conditions for compatibility with additivity, and we present such conditions due to Kraft *et al.*, Luce, and Savage.

As a prelude to Scott's theorem we represent subsets A of Ω by points in a linear vector space $\mathscr{V}$; $\mathscr{V}$ is the linear manifold generated from the set indicator (characteristic) functions

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A; \end{cases}$$

that is,

$$\mathscr{V} = \left\{ v : \exists (c_i \in R^1, A_i \subseteq \Omega)\, v = \sum_i c_i I_{A_i} \right\}.$$

The CP relation $\gtrsim$ between subsets of Ω partially defines a relation $\gtrsim'$ between elements of $\mathscr{V}$ as follows: $A \gtrsim B \Rightarrow I_A \gtrsim' I_B$; $c > 0$, $u \gtrsim' v \Rightarrow cu \gtrsim' cv$; $u \gtrsim' v \Rightarrow u + w \gtrsim' v + w$. While we would also wish to assume that $u \gtrsim' \mathbf{0}$, $v \gtrsim' \mathbf{0} \Rightarrow u + v \gtrsim' \mathbf{0}$, this is not necessarily consistent with $\gtrsim$; to wit, the KPS counterexample. We require an additional hypothesis concerning $\gtrsim$ to guarantee this property for $\gtrsim'$, and one form is provided in the statement of the following.

Theorem (Scott). For finite Ω, the CP relation $\gtrsim$ satisfying C1–C4 is compatible with finite additivity if and only if

$$(\forall u_i, v_i) \left(\sum^n u_i = \sum^n v_i,\ (\forall i < n)\,(u_i \lesssim v_i) \Rightarrow u_n \gtrsim v_n \right). \qquad (**)$$

Proof. The proof is based on the convexity of $\{v : v \geqslant' \mathbf{0}\}$ and the hyperplane separation theorem for disjoint convex sets. See Scott [8, pp. 246–247]. ∎

Extension. Scott claims that use of the Hahn–Banach theorem enables him to extend this theorem to the case of infinite Ω.

3. Should CP Be Compatible with Finite Additivity ?

When $\gtrsim$ satisfies C0–C4 a contradiction to (**) can only occur if $\sum u_i$ is not itself an indicator function. In essence, Scott's theorem shows that $\gtrsim$ is compatible with a finitely additive representation if and only if when $\gtrsim$ is extended to a collection of objects that are no longer events it does not become possible to find inconsistencies in the ordering of the nonevents. While this observation is of interest, it does not provide a reasonable axiom for a theory of CP concerned solely with events. Why should we be concerned about objects that have no reasonable interpretation in terms of random phenomena? Inconsistencies that arise outside the domain of events, as modeled by sets or indicator functions, do not reflect adversely on the adequacy of the theory when restricted, as it would be, to the domain of events. For example, what in our understanding of CP compels us to eliminate the ordering of (*) as a CP relation?

While the relationship between additivity and probability is discussed in greater detail in Section IIID, we should note that from the viewpoint of the theory of measurement it is only reasonable to insist upon an additive scale (probability) for uncertainty if this numerical relationship reflects an underlying empirical relationship between uncertainties. Beyond C4, we are unaware of any such relationship between uncertainties that would urge an additive representation. It might be thought that if the interpretation underlying the formal structure of CP is a relative-frequency one, then we will be inexorably led to consider only those CP relations having an additive representation. However, the falsity of this conjecture is demonstrable through the following example of the estimation of a CP relation from relative-frequency data [1].

Let $N_A(n)$ be the number of occurrences of $A \subseteq \Omega$ in n unlinked repetitions of the random experiment $(\Omega, \mathscr{F}, \gtrsim)$. Conformity between relative-frequency and CP presumably suggests that

$$N_A(n) > N_B(n) \Rightarrow A \gtrsim B.$$

However, especially for small n, we would not feel obliged to accept

$$N_A(n) = N_B(n) \Rightarrow A \approx B.$$

With this understanding of the estimation of CP from relative-frequency, we could estimate (*) from the following data:

$$N_a(n)/n = 1/16, \quad N_b(n)/n = 2/16, \quad N_c(n)/n = 3/16,$$
$$N_d(n)/n = 4/16, \quad N_e(n)/n = 6/16.$$

The indications are that the class of reasonable CP relations is significantly larger than the class of relations compatible with a finitely additive representation. However, there seem to be significant difficulties in carrying out the combinatorial analysis necessary to make this statement precise for finite Ω.

4. Sufficient Conditions for Compatibility with Finite Additivity

Sufficient conditions for compatibility of CP with finite additivity have been suggested by Kraft *et al.*, Savage, and Luce. After introducing these conditions we will show that Luce's suggestion is uniformly best, although the other two initially have greater intuitive appeal. Kraft *et al.* suggest use of the property of polarizability, defined as follows.

Definition. A is polarizable if $(\exists A', A'')(A = A' \cup A'', A' \cap A'' = \varnothing, A' \approx A'')$.

The Kraft *et al.* condition is

K. $(\forall A)$ (A is polarizable).

Savage introduced the notions of fine and tight CP relations but we have shown that his weaker assumption of the existence of almost uniform partitions when combined with C5 will suffice for finite additivity.

Definition. An n-fold almost uniform partition $\mathscr{P}_n = \{P_j^{(n)}\}$ is an n-set partition of Ω with the property that the union of no k sets in $\mathscr{P}_n$ is more probable than the union of any $(k + 1)$ sets in $\mathscr{P}_n$.

Savage then hypothesizes

S. For infinitely many n, there exists $\mathscr{P}_n$.

Finally, Luce [4] proposes

L. $(A \cap B = \varnothing, A \gtrsim C, B \gtrsim D) \Rightarrow (\exists C', D', E)(C' \cap D' = \varnothing,$
$E \approx A \cup B, C' \approx C, D' \approx D, E \supseteq C' \cup D')$.

Clearly conditions K and S presume that Ω is infinite, whereas L is compatible with some finite Ω.

It is evident that condition K implies condition S with $\mathscr{P}_n$ existing for all n of the form 2^j. Hence it suffices to treat S and ignore K.

With regard to S, Savage [9] has proven the following.

Theorem (Savage). If $\gtrsim$ satisfies C1–C4 and S, then there exists a unique (to within multiplication by a positive constant) finitely additive function P such that

$$A \gtrsim B \Rightarrow P(A) \geqslant P(B) \qquad \text{(almost agreement)}.$$

Furthermore,

$$(\forall 0 \leqslant \rho \leqslant 1)(\forall A)(\exists B \subseteq A)\,(P(B) = \rho P(A)). \qquad (***)$$

To close the gap between almost agreement (implication in one direction) and agreement (implication in both directions) we introduce

Theorem 5. If a finitely additive P satisfies (***) and almost agrees with $\gtrsim$ satisfying C0–C4, then P agrees with $\gtrsim$ if and only if $\gtrsim$ satisfies C5.

Remark. In this case the gap between almost and strict agreement is solely due to the possible nonexistence of any (not necessarily additive) agreeing function.

Corollary. If $\gtrsim$ satisfies C0–C5 and S, then there exists a unique (to within multiplication by a positive constant) finitely additive P satisfying (***) and agreeing with $\gtrsim$.

This corollary informs us as to the most that can be expected of those CP relations compatible with the hypothesis of infinitely many almost uniform partitions.

With respect to condition L, Luce [4] has shown the following.

Theorem (Luce). If the CP relation $\gtrsim$ satisfies C0–C4, C6, and L, then there exists a unique (to within multiplication by a positive constant) finitely additive function P agreeing with $\gtrsim$.

To demonstrate that (C0–C4, C6, L) provide a uniformly better characterization of finite additivity than does (C0–C5, S), we establish

Theorem 6. (C0–C5, S) $\Rightarrow$ (C0–C4, C6, L), but the converse implication is not valid.

To date, the set (C0–C4, C6, L) seems to be the best published characterization of the subset of CP relations compatible with finite additivity, where by "best" we refer to the intuitive acceptability of the axioms as well as to the size of the subset. Some insight into C6, L is provided by the discussion in Section IIID of the empirical meaning of additivity for a probability measurement scale. That discussion leads us to conclude

that (C1–C4, C6, L) is too stringent a characterization of those CP relations for which additive P is empirically meaningful.

Results on the related problem of the existence of almost agreeing finitely additive P are available in Fishburn [10] and will not concern us here.

IID. Compatibility with Countable Additivity

In view of the widespread acceptance of the Kolmogorov axioms (see Section IIIA) for probability, it is of interest to characterize the subclass of CP relations that are compatible with countably additive agreeing P. As is well known in probability theory, countable additivity is equivalent to finite additivity and the following continuity condition:

$$B_i \downarrow \varnothing \Rightarrow \lim_{i\to\infty} P(B_i) = 0.$$

The continuity condition suggests the following weakened version of C7 as an axiom for CP.

C8. $(\forall\{B_i\})\left(B_i \downarrow \varnothing \Rightarrow \bigcap_{i=1}^{\infty} \{A : \varnothing \prec A \preccurlyeq B_i\} = \varnothing\right).$

The role of C8 can best be seen from

Theorem 7. If $\gtrsim$ admits of an agreeing finitely additive P, then P is countable additive if and only if $\gtrsim$ satisfies C8.

Corollary. If $\gtrsim$ satisfies C1–C4, C7, and L, then there exists a unique agreeing P satisfying the Kolmogorov axioms.

Villegas [5] has proposed a characterization of a subset of CP relations compatible with countable additivity. After introducing the hypothesis that there are no atoms,

$$\text{(atomless)} \quad (\forall A \succ \varnothing)(\exists B \subset A)\,(A \succ B \succ \varnothing),$$

he then establishes

Theorem (Villegas). If $\gtrsim$ satisfies C0–C4 and C7 and is atomless, then there is a unique agreeing P satisfying the Kolmogorov axioms.

Unfortunately Villegas' simple conditions yield a theory of countable additivity that is strictly less powerful than that of the preceding to complete a probability assignment unnecessarily, and yet it represents a collection of events for which the probability axioms (particularly countable additivity) are well defined and reasonable. In fact the axioms for Λ require us, given events $\{F_i \in \Lambda\}$, to include only those other events whose probabilities are determinable solely from the given $\{P(F_i)\}$ and the Kolmogorov axioms for probability. Furthermore, from a mathematical viewpoint, Λ-fields are monotone classes (monotone sequences of sets which are ordered by inclusion have limits in the class) and would allow discussion of limits. However, Λ is not particularly suited for the definition of conditional probability. It is possible that $F_1, F_2 \in \Lambda$ but $F_1 \cap F_2 \notin \Lambda$; thus we could not define $P(F_2 \mid F_1)$.

The π-field is defined by

$$F, G \in \pi \Rightarrow F \cap G \in \pi,$$

or if we wish to ensure closure under monotone decreasing limits,

$$(\forall i) F_i \in \pi \Rightarrow \bigcap_{i=1}^{\infty} F_j \in \pi.$$

The advantage of the π-field over the Λ-field is that it enables us to define conditional probabilities. A disadvantage of the π-field is that we may have to include sets as events whose probabilities are not readily available.

4. The von Mises Field of Events

The von Mises field $\mathscr{V}$ arises from a careful study of experiments consisting of repeated, unrelated trials, each trial having a finite number of possible outcomes (e.g., repeated coin tossing). One may assume that such experiments and the relative-frequencies of occurrence of the various outcomes lie at the bottom of the Kolmogorov system [1, p. 4]. Hence it is significant that von Mises' investigations [5] led him to value a field of events $\mathscr{V}$ which, in general, differs from a σ-field. The sample space for the outcomes of indefinitely repeated experiments, each experiment having outcomes in Ω, is taken as the set Ω^∞ of all infinite sequences with elements $\omega_i \in \Omega$. All events whose occurrences are determined by the first n experiments (cylinder sets) are in the field $\mathscr{V}$. We then construct the σ-field $\mathscr{F}_n$ of unions and complements of events depending on the first $m \leqslant n$ outcomes, and define $\mathscr{F}' = \bigcup_{n=1}^{\infty} \mathscr{F}_n$; $\mathscr{F}'$ is a field (no longer a σ-field) all of whose elements are finite coordinate cylinder sets. The field $\mathscr{F}'$ can then be extended to $\mathscr{V}$ as follows: $\mathscr{V}$ is the smallest field satisfying

$$(\forall n)(\exists L_n, U_n \in \mathscr{F}_n)(L_n \subseteq V \subseteq U_n)(\lim_{n\to\infty}[P(U_n) - P(L_n)] = 0) \Rightarrow V \in \mathscr{V}.$$

The field $\mathscr{V}$ contains sets measurable in the sense of Peano and Jordan. The events in $\mathscr{V}$ have the important property that their probability can be "measured" to any desired degree of accuracy by repetitions of verifiable events (events whose outcomes are decided by a finite number of experiments). Events not in $\mathscr{V}$ do not have experimentally assessable (by relative-frequencies) probabilities.

If, for example, we take the coin-tossing experiment, then the event $V = \{\text{heads occur only finitely often}\}$ is not in the von Mises field although it is in a σ-field generated from the cylinder sets. The tail event V does not depend on the first n outcomes of the repeated experiment for any finite n. Thus L_n must be the null set and U_n the sample space and $\lim_{n\to\infty}[P(U_n) - P(L_n)] = 1$. Hence the von Mises field $\mathscr{V}$ does not include the minimal σ-field generated from the cylinder sets. It is also true that the minimal σ-field need not include $\mathscr{V}$, and $\mathscr{V}$ is contained in the completion of the minimal σ-field. Von Mises' analysis successfully challenges the axiomatic imposition that the proper collection of events must be a σ-field.

IIIC. Overspecification in the Probability Axioms: View from Comparative Probability

One approach to an understanding of the restrictive and arbitrary aspects of the Kolmogorov axioms for quantitative probability is to start from the weaker and, therefore, less controversial concept of comparative probability and build up to the Kolmogorov concept of probability.

1. Unit Normalization and Nonnegativity Axioms

As can be seen from axiom C5 and Theorem II.1, it is an assumption that uncertainty or chance can be quantitatively described, although one we accept. Even when the CP relation $\gtrsim$ satisfies C1–C7, the agreeing quantitative probability P is not uniquely determined. Any strictly increasing transformation f of P, where $P' = f(P)$, results in an agreeing probability P'. This lack of unicity allows us to choose P such that

$P(\Omega) = 1$, $P(A) \geqslant 0$, thereby satisfying the first two Kolmogorov axioms. However, nothing necessitates this choice; it is simply a convention. The problem with conventions is that we may lose sight of their arbitrary origin. The unit normalization and nonnegativity conventions begin to appear as substantive properties and much statistical and computational effort is expended to ensure that these properties hold for estimates of, and calculations with, probability.

2. Finite Additivity Axiom

As we have commented in Section IIC, we are not convinced that orderings of random events must necessarily admit of finitely additive representations. The assumption that P must be additive is restrictive of the notion of CP; for example, we must rule out cases such as the Kraft *et al.* example (∗). The discussion of Scott's theorem in Section IIC and of the empirical meaning of additivity to be given in Section IIID suggests that the property of additivity is neither logically necessary nor necessarily meaningful.

The rationale underlying the almost universal choice of a finitely additive probability seems to be compounded from considerations of the important relative-frequency interpretation and the intuitive but incorrect conclusion that we require additivity to reflect the properties asserted in the comparative probability axioms C1–C4. As we have previously noted, a relative-frequency-based measure of chance is compatible with a CP relation $\gtrsim$ that admits of no additive agreeing probability; for example, if N_A is the number of occurrences of event A in unlinked repetitions of an experiment, then there are CP relations satisfying

$$N_A > N_B \Rightarrow A \gtrsim B$$

for which no agreeing P is additive. A more general approach to the resolution of the question of the necessity for a CP relation to have an additive agreeing probability is to adopt a rational method for the estimation of CP and then to see if it inexorably leads to additivity. Reflection on the decision theory sketched in Section IIG and its extensions suggests that the particular CP relation that best fits our prior knowledge about some random phenomenon and our available observations will not necessarily be compatible with additivity. A blind insistence upon modeling chance phenomena only through additive probability forces us to ignore potentially preferable descriptions.

Finally, even if we agree to restrict CP to only those relations admitting of additive P, the choice of additivity as distinct, say, from multiplicativity is another arbitrary convention. As discussed in Subsection IIID3, there are infinitely many probability scales equivalent to the additive scale and no significant conceptual basis for preferring one scale to the other.

3. Continuity Axiom

If $\gtrsim$ satisfies C8, then any finitely additive P will be continuous and, by a well-known theorem in probability, also be countably additive. While C8 is not unreasonable, this technical requirement has an appreciable effect in ruling out possible probability assignments. As is well known, it is not always possible to assign arbitrary countably additive probabilities to arbitrary collections of subsets of Ω nor even to arbitrary σ-fields. By the Caratheodory extension theorem [6] we can always uniquely extend a countably additive probability specified over a field to the smallest σ-field containing the field and hence to the completion of the σ-field. In general we cannot extend the assignment of P to other sets, although specific assignments can often be extended somewhat. For example, there is no countably additive measure extendible to all subsets of Ω and assigning intervals in $\Omega = (0, 1)$ their length.

A more interesting example is provided by the experiment of "drawing a positive integer at random." By drawing a positive integer at random we might mean in part that $\Omega = (1, 2, 3, ...)$, $\mathscr{F}$ is the power set of Ω (set of all subsets of Ω), and P assigns zero probability to all finite subsets of $\mathscr{F}$. If P is countably additive, then, by the Caratheodory extension theorem, P extends to $\mathscr{F}$ by being identically zero, in violation of $P(\Omega) = 1$. Thus there is no countably additive measure capable of describing this somewhat idealized experiment. One resolution of this difficulty is provided by Renyi's axiomatization of conditional probability described below. A second resolution is to relax the unit normalization and allow, say, σ-finite measures, although this could violate C4 and C6. A third resolution is to replace countable additivity by finite additivity. If we have only finite additivity, we cannot conclude from $P(\omega_i) = 0$ that $P(\Omega) = 0$. We might define P for those subsets F of Ω for which the following limit (density) exists: Let $N_F(n)$ be the number of elements of F that are less than or equal to n and define $P(F) = \lim_{n\to\infty}(1/n)\, N_F(n)$. This yields a nonnegative, finitely additive measure such that $0 \leqslant P(F) \leqslant P(\Omega) = 1$.

Nor can we find a justification for countable additivity in an unsophisticated appeal to the limit of a relative-frequency interpretation of probability. If $\Omega = \{\omega_i\}$ and $N_{\omega_i}(n)$ is the number of occurrences of ω_i in the first n trials, then it does not follow that

$$\sum_i \lim_{n\to\infty} \frac{N_{\omega_i}(n)}{n} = \lim_{n\to\infty} \sum_i \frac{N_{\omega_i}(n)}{n}.$$

In fact, while

$$\sum_i \frac{N_{\omega_i}(n)}{n} = 1,$$

it is possible for

$$(\forall i)\, P(\omega_i) = \lim_{n\to\infty} \frac{N_{\omega_i}(n)}{n} = 0.$$

IIID. Overspecification in the Probability Axioms: View from Measurement Theory

1. Fundamentals of Measurement Theory

The examination of the relation between comparative and quantitative probability exposes some of the formal conditions that make it possible to choose P satisfying the Kolmogorov axioms. It does not, however, show why that choice is either meaningful or necessary. Choosing the Kolmogorov axioms for P amounts to, at the least, a choice of a measurement scale for the random phenomena of chance and uncertainty. To better appreciate what is involved in choosing a meaningful measurement scale, we digress and follow Suppes' and Zinnes' treatment of measurement theory [7], as also treated in [8].

Fundamental measurements can be thought of as establishing a homomorphism between a given empirical relational system $\mathscr{E}$ and a selected numerical relational system $\mathscr{N}$.

The $\mathscr{E}$ system consists of a domain of real objects or events together with a collection of (significant) relations between elements of the domain; for example, the domain might contain many masses and a relation might be that of "heavier than" as determined by a balance scale. The $\mathscr{N}$ system consists of a domain of real numbers together with selected numerical relations; for example, the domain might be the positive reals and a relation might be that of "greater than." The measurement system is a triple $(\mathscr{E}, \mathscr{N}, f)$ with a mapping f (homomorphism) between the domains of $\mathscr{E}$ and $\mathscr{N}$ that preserves

corresponding empirical and numerical relations; for example, by means of a spring scale we can assign positive numbers to the masses such that "heavier than" between masses corresponds to "greater than" between their numerical weights.

Measurement systems can be classified according to the groups of transformations of f that leave unchanged the correspondences between relations. For example, any strictly increasing function of f would preserve a correspondence between "heavier than" and "greater than"; such a measurement system is called an *ordinal* scale. An *absolute* scale is one for which only the identity transformation preserves the relational homomorphisms. There are few physical concepts that require an absolute scale. The number of objects in a collection requires an absolute scale, but mass, length, time, temperature, current, voltage, etc., do not.

To clarify the measurement status of probability let us first discuss the familiar concept of weight. We assume that we have a set of individual masses $a, b, c, \ldots$. There is an operation of combining or packaging masses together. By an object we mean either an individual mass or a package of masses from our collection. The basic empirical relation between masses is established through a two-pan balance scale. The measurement of weight, ω, as carried out perhaps by a spring scale, assigns positive numbers to objects. If we are only interested in whether an object A is heavier than an object B, then the measurement scale ω need only be such that "A is at least as heavy as B," if and only if, "$\omega(A) \geqslant \omega(B)$." In such an ordinal scale the assertion "$\omega(A)/\omega(B) = k$" has no meaning beyond that of the implicit inequality. For the numerical statement "$\omega(A)/\omega(B) = k$" to be empirically meaningful there must be an empirical relation that is transformed by the function (scale) ω into it.

In order to interpret the numerical ratio statement we might introduce an empirical concept of a standard series of weights modeled on Luce's definition of a standard series given in Section IIB. By a standard series of weights $\{B_i\}$ from object B to object A we mean the following: B_1 balances B; C_j balances B; C_j and B_j contain no common object; B_{j+1} balances the package of C_j and B_j; and for some n, B_n balances A. (Note that we permit an object to "balance" itself.) We can now supply an empirical interpretation for "$\omega(A)/\omega(B) = k$" as follows: There is an object C such that there is a standard series of length n from C to B and of length kn from C to A.

We can also interpret the numerical difference statement "$\omega(A) - \omega(B) = z$" as corresponding to the following empirical relation: There are objects C, D, and E, where C balances B, C and D

have no objects in common, E balances A, the package of C and D balances E, and $\omega(D) = z$.

The preceding constructions do not exhaust the possibilities for establishing an empirically meaningful additive scale. More generally we might conclude that ω is an empirically meaningful additive scale if, to within transformation through multiplication by a positive constant, it is the unique scale satisfying

(a) A is at least as heavy as $B \Leftrightarrow \omega(A) \geqslant \omega(B)$,

(b) $A \cap B = \varnothing \Rightarrow \omega(A \cup B) = \omega(A) + \omega(B)$.

The fundamental ideas of measurement scales are presented in some detail in Krantz *et al.* [8]. If our criterion of empirical meaningfulness is violated, then statements of ratios or of differences are happenstance and need have no empirical significance beyond that of establishing inequality.

2. Probability Measurement Scale

To apply the preceding discussion of weighing directly to probability we correspond elementary outcomes to individual masses, Ω to the collection of masses, events to objects, and the operation set union to packaging masses. The probabilistic analog to the balance scale which will enable us to compare empirically the probabilities of events depends on our interpretation of probability (e.g., relative-frequency of occurrence, willingness to wager on occurrence, etc.). Without involving ourselves at this moment in interpretive questions, let us assume the existence of some mechanism that can determine the CP relation $\succsim$. Probability is then a mapping in a measurement system where the domain of $\mathscr{E}$ is a field of events and the domain of $\mathscr{N}$ is the unit interval [0, 1]. The empirical relations concern the uncertainty, or chance, of the occurrence of an event in the performance of an experiment.

Reasonable empirical interpretations for ratio or difference statements of probabilities follow directly from our discussion of weight. However, if we are unable to find suitable standard series, etc., then the usual probability statements are meaningless numerical accidents. Now, while it is common in thinking of weighing to imagine that we can create a large collection of weights and thereby find the needed standard series, etc., this is not the case in applications of probability. In applications of probability theory we generally deal with a given possibly finite set Ω and field $\mathscr{F}$. It is evident that for finite Ω we cannot generally expect that ratio and difference statements will be meaningful. For

instance, if $\Omega = \{0, 1\}$ with $\varnothing \prec 0 \prec 1 \prec \Omega$, then there is no empirical meaning in $P(1) = 3P(0)$, beyond the information that $P(1) > P(0)$. In fact, it is only when Luce's axiom (L) for CP (Section IIC) is satisfied that an empirical interpretation of ratio and difference statements can be given using standard series. Nor is there any way of satisfying our general criterion for the empirical meaningfulness of an additive scale, except in the case of $P(1) = P(0)$. How can this conclusion be reconciled with the widespread belief in the meaningfulness of Kolmogorov's probability function for finite probability models?

One possibility at reconciliation is that there may be another empirical relation that maps into, and thereby interprets, the usual numerical probability statements. Concerning this possibility we have no considered ideas.

Another possibility is that finite probability models may, in reality, be elliptical references to models with larger or even infinite sample spaces. For example, we may assert that a random experiment is such that $\Omega = \{0, 1\}$, $P(1) = .75$. A relative-frequentist would claim that $P(1)$ was derived by examining the more complex random experiment having as realizations n-tuples of binary outcomes describing independent repetitions of the original experiment. Hence the probability statement originates with an experiment having a sample space with 2^n points, rather than 2 points, in which it is meaningful.

The crux of this explanation can be put in a form that does not depend on a particular interpretation of the probability concept. Savage and Kraft *et al.* in their polarizability axiom (see Section IIC) have considered the possibility of establishing the existence of an additive P representation for a random experiment by requiring that the CP relation $\succsim$ be extendible in a consistent manner to the compound experiment that augments the given experiment with a coin-tossing experiment. This device can also be used to interpret the additive P scale. Let $\Omega = \{\omega\}$ and $\succsim$ describe a random experiment. Let $\Omega_n = \{\omega_n\}$ be a sample space for n "independent" tosses of a "fair" coin; that is, in whatever interpretation we have for probability, the tosses are such that all outcomes are equally probable, $\omega_n \approx \omega_n'$. The compound experiment has a sample space $\Omega^* = \{(\omega, \omega_n)\}$ and an ordering $\succsim^*$. The ordering $\succsim^*$ is consistent with $\succsim$ in that

$$(\omega, \omega_n) \succsim^* (\omega', \omega_n') \Leftrightarrow \omega \succsim \omega'.$$

It may now be possible to make meaningful numerical statements about the compound experiment that were not possible in the original experiment. For example, we can meaningfully claim that $P(1) = 3P(0)$, or $P(1) = .75$, when $\Omega = \{0, 1\}$ by referring it to the empirical proposition

$$(1, HH) \approx^* (0, TT) \cup (0, TH) \cup (0, HT).$$

The explanation above is not an argument for the general acceptability of the polarizability axiom or for a freedom to always associate an independent, fair coin-tossing experiment with any given random experiment. This would put the cart before the horse. We have argued that the usual finite probability model $(\Omega, \mathscr{F}, P)$ may suppress the description of the more complex finite, or infinite, probability model that provided the information on P. The introduction of the coin-tossing experiment is just a formal device to generate a canonical reconstruction of the complex, original random experiment. In the more realistic finite Ω models of random experiments, using CP in place of quantitative probability, there may never have been a more complex "parental" experiment. In this case the canonical reconstruction, even when possible, is not compelling in its conclusions; we cannot create meaning out of nothingness.†

Unlike the case of mass, length, and so forth, we do not believe that the tendency for an event to occur can always be split into equal pieces.

It appears that the Kolmogorov probability scale may be unnatural when we deal with random experiments having finitely many outcomes. Not only may an additive scale not exist (e.g., the counterexample of Kraft *et al.* in Section IIC), but even when it exists it may be empirically meaningless. The inadequacy of the Kolmogorov axioms, from this point of view, appears less acute when we deal with "reasonable" infinite sample space models.

3. Necessity for an Additive Probability Scale

Even if we are able to interpret the Kolmogorov absolute probability scale meaningfully we have yet to argue for the necessity of this scale. In fact, there are many equivalent scales that are not additive. From

† An analogy may clarify this point. The reader, by invoking his knowledge of vocabulary and grammar, understands the statement, "It's a beautiful day, and I feel fine." However, if he now learns that this statement appears in a long list generated by a randomly programmed computer, would he still think it appropriate to evaluate the meaning of the statement by bringing the same language tools to bear? In the view we have been proposing, the statement has meaning in the usual case because we assume that it was made by a person. Learning that a computer generated it does not change the statement but radically changes its meaning.

Theorem II.4, we know that, under reasonable axioms for comparative probability, there exist P, G, and

$$A \cap B = \varnothing \Rightarrow P(A \cup B) = G(P(A), P(B)).$$

It can be shown [9] that under mild regularity conditions the strictly increasing, symmetric, associative function G can be represented in the quasilinear form

$$G(x, y) = h(h^{-1}(x) + h^{-1}(y)),$$

where h is some strictly increasing function. From the quantitative probability P we can derive an additive probability P' through the definition $P'(F) = h^{-1}(P(F))$. The function G' corresponding to P' is then $G'(x, y) = x + y$. What real considerations though make us prefer the particular choice of measurement scale h^{-1} yielding an additive, quantitative probability P'?

If we think of probability as a model of relative-frequency, then we seem to be led to accept the choice of $G(x, y) = x + y$. However, agreement on probability as built upon relative-frequency as the quantitative core of chance phenomena does not eliminate the possibility that P may depend on some function of the relative-frequency other than the identity function. For example, we might interpret $P(F)$ as $\lim_{n\to\infty}(N_F(n)/n)^2$, where $N_F(n)$ is the number of occurrences of the event F in n trials. In this case we would find that F_1 and F_2 disjoint imply that $P(F_1 \cup F_2) = [\sqrt{P(F_1)} + \sqrt{P(F_2)}]^2$, the scale transformation h^{-1} taking P into an additive probability assignment being $h^{-1}(x) = \sqrt{x}$. If we accept the fact that there is exactly as much information about a chance event F in $(N_F(n)/n)^2$ as in $N_F(n)/n$, then it becomes only a matter of habit or convenience as to which of these we prefer to work with—a matter of choice and not necessity. (But see Section IVJ.)

An analogy to this situation is that of coordinate-free results in analytic and vector geometry. The choice of an additive scale is analogous to choosing a fixed coordinate system and expressing all results in that system, although the results do not depend on the particular system chosen. The danger in this procedure when applied to probability is the ease with which we forget that the results are coordinate-free and then attempt to make something of the choice of system. The Kolmogorov scale for probability is in large part a convention and occasionally its use evokes empirically meaningless statements.

In sum the observations of Sections IIIC and IIID suggest that the Kolmogorov axioms for probability rule out of consideration possible orderings of random events and then adopt conventions with regard to the remaining orderings of events. The conventions make the choice of probability scale spuriously specific and make attempts to extract significant probabilistic relations from data more difficult. The user finds himself forced to engage in arbitrary assumptions, to match those of the theory, which appear naked in the light of a real application.

IIIE. Further Specification of the Event Field and Probability Measure

1. Preface

Thus far our remarks have indicated that the widely used Kolmogorov setup may be unnecessarily restrictive in certain applications. However, it can also be argued that this theory needs additional axioms before it can be reasonably and profitably employed to model random phenomena; the axioms say little and some of what they do say is either unimportant or a matter of convention. There are several points at which the Kolmogorov axiomatic theory needs to be augmented and discussed.

(1) In applications one generally chooses a collection of events and a family of possible probability assignments, prior to any analysis. What are the guidelines for these choices?

(2) Should the concepts of stochastic independence and conditional probability enter probability theory merely as definitions or do they warrant a presentation coordinate with, or even prior to, that of absolute probability?

(3) What distinguishes probability and its domain from, say, length or mass?

We only hope to clarify, and not answer, these questions in the remainder of this chapter. The lack of an adequate set of answers will partially motivate us to consider, in subsequent chapters, other approaches to probability.

2. Selecting the Event Field

Guidelines for the selection of an event field have been offered in the important cases of experiments with outcomes describable by either finitely, countably infinitely, or uncountably many random variables. The σ-field usually assumed when we deal with the collection of events for a real random variable is the Borel field of subsets of the real line; Cramer [10] has gone so far as to include this suggestion

in the axioms for probability. The Borel field is the smallest, in the sense of inclusion, σ-field containing all intervals; the field may then be completed for a given probability measure by including sets that differ from sets in the Borel field by sets of zero probability. The Borel field is generally thought to be sufficiently rich in events for most problems. Furthermore, it is felt that since the Borel family of events is generated from the simple and basic events of the experimental outcomes being greater than or less than a given number, they are particularly appropriate for the description of experiments.

The Borel field is also the usual choice when dealing with sequences of random variables, for example, repeated experiments with real-valued outcomes. The Borel field of events for an infinite sequence of random variables is the smallest σ-field containing, for every finite n, all of the cylinder sets with n-dimensional intervals as bases. In modeling random processes we encounter the most general case of collections of random variables. It has been suggested that in such instances the Borel field of events be restricted to be standard, that is, the event field should be point isomorphic to some Borel field of subsets of the line [11]. Whatever the adequacy and naturalness of the Borel field for finitely many random variables, von Mises would challenge its use for an infinite sequence of random variables.

An important restriction on the von Mises field $\mathscr{V}$, absent from the definition of σ-field, is that of "conceptual verifiability." An event for an experiment with indefinitely repeated trials is considered to be conceptually verifiable if it is either determined by a finite number of trials or can be approximated by events whose outcomes are determined by a finite number of trials. Von Mises restricts the class of events to those whose probability can conceivably be estimated from the relative-frequency of their occurrence. If an event cannot be approximated by events whose outcome is determined in a finite time, then it is unreasonable to think of repetitions of this event. While von Mises' position is a reflection of his study of single sequences of repeated trials, recourse to an ensemble of repeated experiments does not remove the difficulty. It will still take an infinite time before we gain even an approximate idea of the relative-frequency of occurrence of an event not in $\mathscr{V}$.

3. Selecting P

The desire for a more structured theory of probability is manifest in a suggestion of Gnedenko and Kolmogorov [12] for the modeling of a nondenumerable collection of random variables (random process).