67 On Choosing a Function at Random

1 Introduction

We use the prefix "*m*-" to abbreviate "measurable." Let X and Y be *m*-spaces[1] and let Y^X denote the set of all *m*-transformations from X into Y. We are interested in random procedures for choosing a member of Y^X. An example of such a random procedure is a stochastic process; here X is the time axis. Another example is a mixed strategy in a game, in which the player has to pick a member of Y on the basis of information which may vary over X (here Y^X is the set of pure strategies).

The relation of the concepts we use here to standard concepts from the theory of stochastic processes will be briefly explored in section 7.

2 Distributions over Function Space

There are two approaches to the problem of formalizing the intuitive notion of "random procedure for choosing a member of Y^X." First, we may define a distribution—i.e. probability measure—on Y^X. However, since Y^X is not endowed with any *m*-structure to begin with, we must impose an *m*-structure on it as part of the definition of "distribution." Thus a distribution on Y^X should be defined as a pair $(\mathscr{C}, \mu)$, where $\mathscr{C}$ is a σ-ring on Y^X and μ is a measure on $\mathscr{C}$.

Let us now choose a function f at random from Y^X, and a point x at random from X, and inquire as to the distribution of a point on a (not necessarily continuous) path, when both path and time are chosen at random. More precisely, let us fix distributions $(\mathscr{C}, \mu)$ on Y^X and ν on[2] X, and let $\varphi : Y^X \times X \to Y$ be the mapping defined by $\varphi(f, x) = f(x)$; we wish to regard φ as a random variable, and seek its distribution. Now φ has a distribution only if it is an *m*-transformation; indeed it is easily seen that for $B \subset Y$, we have $\text{Prob}\{f(x) \in B\} = (\mu \times \nu)\{\varphi^{-1}(B)\}$, so that $\varphi^{-1}(B)$ must be measurable in $Y^X \times X$ whenever B is measurable in Y. This places a restriction on the choice of the *m*-structure $\mathscr{C}$ on Y^X; it should be chosen so that φ is an *m*-transformation.

Rather surprisingly, it is in general *impossible* to define an *m*-structure on Y^X that will satisfy this condition. For example, it is impossible even

This chapter originally appeared in *Ergodic Theory*, edited by F. B. Wright, pp. 1–20.

1. Spaces on which there is defined a σ-ring of *m*-sets but not necessarily a measure. The σ-ring of *m*-sets is called the *m-structure* of the space.

2. Note that there is no need to specify an *m*-structure on X as part of the definition of distribution, as X is provided with an *m*-structure to begin with.

when X and Y are both copies of the unit interval with the usual Borel structure [2] (henceforth this m-space will be denoted I). Put in other words, we cannot define a probability distribution over *all* of Y^X that will enjoy reasonable properties. The question now arises whether we could not achieve our aim by restricting attention to a subset F of Y^X (such as the set of all continuous functions when X and Y are copies of the unit interval). More precisely, for $F \subset Y^X$ define $\varphi_F: F \times X \to Y$ by $\varphi_F(f, x) = f(x)$. Is it possible to impose an m-structure on F so that φ_F will be an m-transformation?

Obviously the answer depends on F; for $F = Y^X$ we have seen that it is negative, whereas when F has only one element it is trivially positive. We will say that an F for which the answer to the above question is positive is *admissible*; the appropriate m-structures will also be called *admissible*. Thus the admissible subsets of Y^X are precisely those over which a probability distribution can reasonably be defined. The problem of characterizing the admissible subsets of Y^X has been solved under fairly wide conditions on Y and X [1,2]. In particular, when Y and X are copies of I, then F is admissible if and only if it is a subset of some Baire class (of arbitrary finite or denumerable order). Thus for example the set of all continuous functions is admissible, as is the set of all functions with discontinuities of the first kind only, and so on.

Returning to our original problem, we see that the correct definition of a distribution over Y^X is a *triple* $(F, \mathscr{F}, \mu)$, where $\mathscr{F}$ is an admissible m-structure on the admissible set F, and μ is a probability measure on $\mathscr{F}$. Intuitively, this is a random procedure for picking a member of Y^X, under which the members of Y^X that "can actually occur" are precisely the members of F.

3 Random Variables over Function Space

In contrast to the first approach, which is based on the idea of a "distribution" over Y^X, the second approach is based on the idea of a "random variable" with values in Y^X. Specifically, let $\Omega = (\Omega, \beta, \lambda)$ be an arbitrary probability space which will serve as our sample space. Intuitively, our "random variable" is a function Θ from Ω to Y^X. Here again we are interested in the distribution of $f(x)$, when both f and x are chosen at random; we may expect that some condition must be placed on Θ to ensure that $f(x)$ *has* a distribution. Fortunately, the appropriate condition is *not* that Θ be an m-transformation, because this would again involve defining an m-structure on Y^X. To state the correct condition, we recall that to every function from Ω to Y^X there is a corresponding function from $\Omega \times X$ to Y; to $\Theta: \Omega \to Y^X$ there corresponds the function $\vartheta: \Omega \times X \to Y$ defined by $\vartheta(\omega, x) = \Theta(\omega)(x)$. It is easily seen that if $x \in X$ is chosen according to a distribution ν, then for $B \subset Y$, we have $\text{Prob}\{f(x) \in B\} = (\lambda \times \nu)\{\vartheta^{-1}(B)\}$. Thus the correct condition is that the function ϑ be an m-transformation. Clearly the simplest procedure—which we adopt—is to define a *random variable* varying over Y^X with sample space Ω to be an m-transformation ϑ from $\Omega \times X$ into Y.

4 Ranges and Admissible Sets

The purpose of this note is to investigate the relation between the concepts of "random variable" and "distribution" defined in the previous sections. Let us define the *range* of a random variable ϑ to be the set of all functions in Y^X of the form $\vartheta(\omega, \cdot)$, where $\omega \in \Omega$. The range is the set of points in Y^X that "can actually occur" under ϑ; thus it corresponds to the notion of an admissible set under the "distribution" definition. Our first question is to what extent this correspondence can be made precise.

THEOREM 1 *Every range is admissible. Conversely, for every admissible set F there is a sample space Ω and a random variable ϑ such that the range of ϑ is F.*

Proof The converse is trivial, because we may take $\Omega = (F, \mathscr{F})$ and $\vartheta = \varphi_F$, where $\mathscr{F}$ is an admissible m-structure on F. For the first part, let ϑ be a random variable with range R. As in the previous section, denote $\vartheta(\omega, \cdot)$ by $\Theta(\omega)$; thus $\Theta: \Omega \to R \subset Y^X$. For every $f \in R$ choose one member ω of Ω, such that $\Theta(\omega) = f$; let Ω' be the subspace of Ω obtained in this way, with the subspace structure (a set is measurable in Ω' if and only if it is the intersection of Ω' with an m-set in Ω). Let ϑ' be the restriction of ϑ to $\Omega' \times X$. Now the restriction of an m-transformation to a subspace is still an m-transformation; hence if we give $\Omega' \times X$ the subspace structure (i.e. as a subspace of $\Omega \times X$), then ϑ' will be an m-transformation. But it is easily verified that the subspace structure on $\Omega' \times X$ is the same as the product structure; hence ϑ' is an m-transformation also when $\Omega' \times X$ has the product structure.

Ω' and R are in one–one correspondence under the correspondence $\omega \leftrightarrow \Theta(\omega)$. Let us impose on R the structure corresponding to that of Ω'; then Ω' and R are isomorphic. Hence $\Omega' \times X$ and $R \times X$ are also isomorphic. Let us denote the isomorphism by $\zeta: R \times X \to \Omega' \times X$; we have $\zeta(\Theta(\omega), x) = (\omega, x)$, where on the right side ω is uniquely defined because of the definition of Ω'. Now $\varphi_R(\Theta(\omega), x) = \Theta(\omega)(x) = \vartheta(\omega, x) = \vartheta'(\omega, x) = \vartheta'\zeta(\Theta(\omega), x)$; thus $\varphi_R = \vartheta'\zeta$. But both ϑ' and ζ are

m-transformations, and therefore φ_R is also an m-transformation. Therefore R is admissible, and the proof of Theorem 1 is complete.

5 Ranges and Admissible Sets when the Sample Space Is Standard

Up to now we have allowed the sample space Ω to be arbitrary; in proving that every admissible set F is the range of some random variable, we even allowed Ω to vary with F. In some applications, though, Ω is a copy of the unit interval I with Lebesgue measure, or can be taken as such without loss of generality. This restriction does not change the fact that every range is admissible, but it does cast doubt on the converse.

Let us further restrict our consideration to the case $X = Y = I$. (This restriction is not as severe as it may at first seem; according to a theorem of Mackey [6], every Borel subset of every separable metric topological space, when endowed with the subspace structure, is isomorphic[3] to I.) As we remarked above, there is in this case an elegant characterization of admissible sets, namely as arbitrary subsets of Baire classes of arbitrary order. We seek now a characterization of ranges.

THEOREM 2 *If X, Y, and Ω are copies[4] of I, then every range is a subset of some Baire class, and every Baire class is a subset of some range.*[5]

This theorem does not give a complete characterization of ranges, similar to the complete characterization of admissible sets mentioned above. For example, I do not even know whether every Baire class is a range; on the other hand, it is highly likely that there exist subsets of Baire classes that are not ranges.[6] What Theorem 2 does do is give an "order of magnitude" characterization for ranges; a range can be as large as a Baire class of arbitrarily high (denumerable) order, but no larger.

Proof That every range is a subset of some Baire class is a trivial consequence of Theorem 1 and the fact that every admissible set is a subset of some Baire class (proved in [2]). To prove the second part of Theorem 2, we define a transfinite sequence $\{F_\alpha\}$, where α ranges over all denu-

3. Two m-spaces are *isomorphic* if there is a one–one correspondence between them, which carries m-sets onto m-sets in both directions.

4. Strictly speaking, Ω is a copy of (I, λ) rather than of I. However we are dealing with measurability properties rather than with measure properties, and will henceforth (for the remainder of the section) ignore the measure on Ω.

5. "Baire class" is a topological concept, so its appearance in a theorem that deals with m-structures should be explained. Since X and Y are in one–one correspondence with I, we can impose on them topologies corresponding to the standard topology on I; the theorem pertains to these topologies. For an intrinsic characterization of Baire classes in terms of m-structures, see [1, 2].

6. The entire discussion is under the assumption $\Omega = X = Y = I$.

merable ordinals, inductively as follows: F_0 is the set of all continuous functions from X to Y; F_α is the set of all functions that are pointwise upper limits of sequences of functions in $\bigcup_{\beta<\alpha} F_\beta$. We will prove by induction on α that F_α is a range; since the α'th Baire class is clearly a subset of F_α, this will complete our proof. To start the induction, let $F_0(a, B) = \{f \in F_0 : f(a) \in B\}$, where $a \in X$ and $B \subset Y$. Let $\mathscr{F}_0$ be the m-structure on F_0 that is generated by all sets of the form $\mathscr{F}_0(r, U)$, where r is rational and U is open. (It happens that $\mathscr{F}_0$ is also generated by the uniform convergence topology on F_0, but this does not concern us here.)

LEMMA 1 *$(F_0, \mathscr{F}_0)$ is an admissible pair.*

Proof We first show that if a is arbitrary and B is measurable, then $F_0(a, B) \in \mathscr{F}_0$. Indeed, let $r_1, r_2, \dots$ be a sequence of rationals converging to a. Then for open $U, f(a) \in U$ if and only if $(\exists N)(\forall n \geqslant N)(f(r_n) \in U)$; thus $F_0(a, U) = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} F_0(r_n, U)$, and hence $F_0(a, U) \in \mathscr{F}_0$. Since the m-structure of Y is generated by the open sets, and the mapping $B \to F_0(a, B)$ carries unions into unions and complements into complements, it follows that $F_0(a, B) \in \mathscr{F}_0$ for all a and all measurable B. The lemma now follows from theorem G of [2].

LEMMA 2 *$(F_0, \mathscr{F}_0)$ is isomorphic to I.*

Proof Denote the infinite product $I \times I \times \dots$ by I^∞; the members of I^∞ are sequences $t = \{t_1, t_2, \dots\}$. The m-structure of I^∞ is generated by cylinder sets of the form $I \times \dots \times I \times U \times I \times \dots$, i.e. by the sets $\{t : t_i \in U\}$, where U is open; such a set will be denoted $I^\infty(i, U)$.

Let $\{r_1, r_2, \dots\}$ be the set of all rationals in I. Define a mapping $\xi : F_0 \to I^\infty$ by $\xi(f) = \{f(r_1), f(r_2), \dots\}$. We first show

i. *ξ is an isomorphism between F_0 and its image $\xi(F_0)$.* Since a continuous function is completely determined by its values on the rationals, $\xi(f_1) = \xi(f_2)$ implies $f_1 = f_2$; in other words, F_0 and its image $\xi(F_0)$ are in one–one correspondence under ξ. Moreover, if $\xi(F_0)$ is considered a subspace of I^∞ with the subspace structure, then ξ maps the generator $F_0(r_i, U)$ of $\mathscr{F}_0$ onto the generator $\xi(F_0) \cap I^\infty(i, U)$ of the structure of $\xi(F_0)$. Hence the generators are also in one–one correspondence under ξ, and a fortiori ξ is an isomorphism between F_0 and $\xi(F_0)$.

Next, we show

ii. *$\xi(F_0)$ is an m-set in I^∞.* This follows from the remark that a function f on the rationals between 0 and 1 can be extended to a continuous function on I if and only if it is uniformly continuous on the rationals. Thus we have $t \in \xi(F_0)$ if and only if

$$(\forall k)(\exists j)(|t_m - t_n| < 1/k \quad \text{whenever} \quad |r_m - r_n| < 1/j).$$

In other words,

$$\xi(F_0) = \bigcap_{k=1}^{\infty} \bigcup_{j=1}^{\infty} \bigcap_{(m,n) \in A(j)} \{t\colon |t_m - t_n| < 1/k\},$$

where $A(j) = \{(p,q)\colon |r_p - r_q| < 1/j\}$. Since it is well known (and easily proved) that the set in curly brackets is an m-set in I^∞, this demonstrates (ii).

I^∞ is known to be isomorphic to I; from (ii) it follows that $\xi(F_0)$ passes under this isomorphism to a Borel subset of I. According to a theorem of Mackey [6], every non-countable Borel subspace B of I is isomorphic to I. Hence $\xi(F_0)$ is isomorphic to I. The lemma now follows from (i).[7]

Lemma 2 says that Ω and F_0 are isomorphic; we may therefore assume without loss of generality that they are identical, and define $\vartheta = \varphi_{F_0}$. Then by Lemma 1, ϑ is an m-transformation, and by Lemma 2, its range is F_0. This starts our induction.

For the inductive step, let α be a finite or denumerable ordinal, and suppose it has been shown that F_β is a range for all $\beta < \alpha$. Let $\Omega_1, \Omega_2, \ldots$ be a sequence of copies of Ω, and let $\vartheta_1\colon \Omega_1 \times X \to Y$, $\vartheta_2\colon \Omega_2 \times X \to Y, \ldots$ be a sequence of random variables such that each F_α with $\alpha < \beta$ is the range of infinitely many of the ϑ_i. The infinite product $\Omega_1 \times \Omega_2 \times \ldots$ is isomorphic to Ω, and as before we suppose without loss of generality that it equals Ω; thus for $\omega \in \Omega$, we may write $\omega = \{\omega_1, \omega_2, \ldots\}$. Define $\vartheta\colon \Omega \times X \to Y$ by $\vartheta(\omega, x) = \limsup_{i \to \infty} \vartheta_i(\omega_i, x)$. From the fact that the ϑ_i are m-transformations, it follows that ϑ is an m-transformation; furthermore it may be seen that the range of ϑ is exactly F_α. This completes the proof of Theorem 2.

Note that we started our induction by showing that F_0 is a range, but it would have been sufficient to show that F_0 is a *subset* of some range; this is in fact easier. We chose to show that F_0 is a range, because this lemma is of interest in itself.

6 The Distribution of a Random Variable

Let $\vartheta : \Omega \times X \to Y$ be a random variable; we wish to define the concept of "the distribution of ϑ". According to section 2, this must be of the

7. I am grateful to B. Peleg for pointing out an error in the original proof of this lemma.

form $(F, \mathscr{F}, \mu)$, where F and $\mathscr{F}$ are admissible and μ is a measure on $\mathscr{F}$. The natural definitions would be as follows: F is the range of ϑ. The m-structure is the *identification* structure; that is, a subset G of F is in $\mathscr{F}$ if and only if $\Theta^{-1}(G)$ is an m-set in Ω. The measure μ is defined by $\mu(G) = \lambda\Theta^{-1}(G)$, where λ is the measure on Ω. These definitions are "natural" in the sense that the induced distribution on Y is the same if ϑ or if $(F, \mathscr{F}, \mu)$ is used.

The only trouble with this "natural" definition is that I do not know whether as defined, the structure $\mathscr{F}$ is admissible.[8] Indeed, let $\imath\colon X \to X$ be the identity, and define $\Theta \times \imath\colon \Omega \times X \to F \times X$ by $(\Theta \times \imath)(\omega, x) = (\Theta(\omega), x)$. Then $\vartheta = \varphi_F \circ (\Theta \times \imath)$, and hence $\vartheta^{-1} = (\Theta \times \imath)^{-1}\varphi_F^{-1}$. Now let B be an m-subset of Y. Then $\vartheta^{-1}(B)$ is an m-subset of $\Omega \times X$, and hence $(\Theta \times \imath)^{-1}\varphi_F^{-1}(B)$ also is. We know that $(F, \mathscr{F})$ is an identification space of Ω under the identification map Θ. If we only knew that $(F, \mathscr{F}) \times X$ is an identification space of $\Omega \times X$ under $\Theta \times \imath$, then the measurability of $\varphi_F^{-1}(B)$ would follow from that of $(\Theta \times \imath)^{-1}\varphi_F^{-1}(B)$, and we could deduce that $\mathscr{F}$ is admissible. The proposition that "if Θ is an identification map and $\imath$ an identity map, then $\Theta \times \imath$ is also an identification map" is intuitively very compelling, but unfortunately I have not succeeded in proving it.[9] Let us call this proposition the "identification space hypothesis"; only the following special cases are known to me:

MACKEY'S RESULT[10] *The identification space hypothesis holds if the domains and images of both Θ and $\imath$ are analytic*[11] *m-spaces.*

ERNEST'S RESULT[12] *The identification space hypothesis holds if Θ carries m-sets onto m-sets.*

Though the hypotheses of both these theorems are quite general, I do not know whether they hold in the situation under consideration, even for some of the simplest cases, (e.g. when $X = Y = \Omega = I$, and ϑ is the random variable with range F_1 that we defined in the previous section).

Our conclusion is that for all we know at the present, a given random variable may have no distribution.

8. By Theorem 1, the set F must be admissible. But the choice of the structure in the proof of that theorem is not unique, and so cannot be used for the current purpose. Even if we arbitrarily pick one of the structures that fit that proof, the resulting distribution may violate the "naturalness" condition of the previous paragraph.

9. The topological analogue is false; the counter-example is an adaptation of an example in Kelley's book [4, p. 132, example G].

10. Private correspondence with Professor G. W. Mackey.

11. I.e., isomorphic with analytic subspaces of I; cf. [6].

12. Private correspondence with Professor J. Ernest.

7 Relation to Other Concepts

What we call a "random variable with values in Y^X" is called a "measurable random function" by Loève [5, p. 502]. A somewhat similar object is called a "measurable stochastic process" by Doob [3, p. 60]; however, Doob fixes the m-structure of X to be the class of all Lebesgue measurable sets. Our "distributions" are closely related to what Doob [3, p. 67] calls a "process of function space type" (i.e. a process in which the sample space is the function space itself). The differences are that Doob considers the set of *all* functions from X into Y, whereas we consider only m-transformations; and Doob imposes a fixed m-structure on function space, namely that generated by all sets of the form $F(a, B) = \{f: f(a) \in B\}$ (of course without regard to admissibility).

What distinguishes the problem discussed here from those of much of the theory of stochastic processes is that we allow x as well as f to vary at random, and inquire as to the distribution of $f(x)$ as a function of both f and x. This makes simultaneous measurability in both variables essential. In stochastic processes one is also interested in the distribution of $f(x)$; but usually only f varies at random, and attention is fixed on some finite x-set. Simultaneous measurability in both variables is then often useful, but not essential.

8 Open Questions

i. Characterize ranges of random variables when $X = Y = \Omega = I$.

ii. In particular, is every Baire class a range in this case?

iii. Prove or disprove the identification space hypothesis.

iv. Does every random variable have a distribution (in the sense of section 6)?

References

1. R. J. Aumann, Spaces of measurable transformations, *Bull. Amer. Math. Soc.* 66 (1960), pp. 301–304 [Chapter 65].
2. R. J. Aumann, Borel structures for function spaces, *Ill. Jour. of Math.* 5 (1961), pp. 614–630 [Chapter 66].
3. J. L. Doob, *Stochastic Processes*, New York, John Wiley (1953).
4. J. L. Kelley, *General Topology*, Princeton, Van Nostrand (1955).
5. M. Loève, *Probability Theory* (second edition), Princeton, Van Nostrand (1960).
6. G. W. Mackey, Borel structures in groups and their duals, *Trans. Amer. Math. Soc.* 85 (1957), pp. 134–165.