Part A

ELEMENTS

I

Probability Spaces

Probability is, with the possible exception of Euclidean geometry, the most intuitive of the mathematical sciences. Chance and its language pervades our common experience. We speak of the chances of the weather turning, getting a raise, electing a candidate to political office, or a bill being passed; we bet on sporting contests in office pools, toss a coin at the start of a game to determine sides, wager on the sex of a newborn, and take a chance on the institutionalised gambling that masquerades as state-sponsored lotteries. And, of course, games of chance have an ancient and honoured history. Excavations of bone dice in archaeological digs in North Africa show that dicing was not unknown in ancient Egypt, board games in which players take turns determined by the roll of dice and card games of some antiquity are still popular in the age of the internet, and the horse race survives as an institution. While the historical palette is rich and applications pervasive, the development of a satisfactory mathematical theory of the subject is of relatively recent vintage, dating only to the last century. This theory and the rich applications that it has spawned are the subject of this book. C ***1–7*** *8, 9*

1 From early beginnings to a model theory

The early history of probability was concerned primarily with the calculation of numerical probabilities for outcomes of games of chance. Perhaps the first book written along these lines was by the eccentric Gerolamo Cardano, a noted gambler, scholar, and bon vivant; his book *Liber de Ludo Aleæ* (Book on Games of Chance) was written perhaps in the 1560s but only published posthumously in 1663.[1] Numerical calculations continued to dominate over the next two and

[1]The modern reader will find Cardano's exhortations have weathered well: "The most fundamental principle of all in gambling is simply equal conditions ... of money, of situation ... and of the dice itself. To the extent to which you depart from that equality, if it is in your opponent's favour, you are a fool, and if in your own, you are unjust." This excerpt is from a translation of *Liber de Ludo Aleæ* by Sydney Henry Gould which appears as an appendix in O. Ore, *Cardano, the Gambling Scholar*, Princeton University Press, Princeton, NJ, 1953.

a half centuries awaiting the birth of a theory but the spread of applications continued unabated until, in the modern day, scarce an area of investigation is left untouched by probabilistic considerations.

Today the informed reader encounters probabilistic settings at every turn in divers applications. The following is a representative list of examples, in no particular order, that the reader will find familiar. (i) The conduct of opinion polls—and what the results say about the population as a whole. (ii) Sampling to determine the impact of an invasive species—or of pollutant concentrations. (iii) The prediction of user preferences—for movies or books or soap—from sporadic internet use. (iv) The search for order and predictability in the chaos of financial markets—or of sunspot activity. (v) Robot navigation over uncertain terrain. (vi) The analysis of noise in communications. (vii) The 3K background cosmic radiation and what it portends for the universe. (viii) The statistical physics of radioactive decay. (ix) The description of flocking behaviour in wild geese and fish. (x) The analysis of risk in the design of actuarial tables. (xi) Mendel's theory of heredity. (xii) Genetic combination and recombination, mutation. (xiii) The spread of infection. (xiv) Estimations of time to failure of machinery—or bridges or aeroplanes. (xv) Investment strategies and probabilities of ruin. (xvi) Queues—of telephone calls at an exchange, data packets at an internet server, or cars in a highway system. (xvii) The statistical search for the Higgs boson and other subatomic particles. The reader will be readily able to add to the list from her common experience.

Following the fumbling early beginnings, inevitably numerical, of investigations of the science of chance, as discoveries and applications gathered pace it became more and more necessary that the mathematical foundations of the subject be clearly articulated so that the numerical calculations, especially in areas that could not be readily tested, could be placed on firm mathematical ground. What should the goals of such a theory be? Given the vast realm of applicability we must hold fast against the temptation to hew the theory too close to any particular application. This is much as in how to reach its full flowering geometry had to be sundered from its physical roots. The rôle and importance of abstraction is in the extraction of the logical axiomatic content of the problem divorced of extraneous features that are not relevant, and indeed obfuscate, to provide a general-purpose tool that can be deployed to discover hitherto unsuspected patterns and new directions. Such a clean axiomatic programme was laid out by Andrei Kolmogorov in 1933.[2]

The key feature of the axiomatic approach is in beginning with a *model* of an idealised *gedanken* chance experiment (that is to say, a thought experiment which may never actually be performed but can be conceived of as being performed). The model, to be sure, makes compromises and introduces convenient mathematical fictions, emphasising certain features of the problem while de-

[2]English speakers had to wait till 1956 for a translation: A. N. Kolmogorov, *Foundations of the Theory of Probability*, Chelsea, New York.

emphasising or ignoring others, both to permit a clear and unobstructed view of essential features as well as to permit ease of calculation. Thus, for instance, we make the pretence that a coin-tossing game persists indefinitely or that a gambler plays with infinite resources; in the same vein, actuarial tables of lifespans permit aging without bound—albeit with incredibly small probabilities—noise waveforms are modelled as lasting for infinite future time, and so on.

In its insistence on a model for the phenomenon under investigation as a starting point the axiomatic theory makes a clean break with the inchoate idea of intuitive probability that we resort to in our daily experience. The classical wager of Laplace that the sun will rise tomorrow, for instance, has no place in the theory abeyant a reasonable model of a chance experiment (that one can conceive of being in repeated use); similar objections hold for assigning chances to doomsday predictions of, say, terrorist nuclear attacks on major cities, or the destruction of earth by a meteor, or to assigning chances for the discovery of some hitherto unknown propulsion mechanism, and so on. This is unfortunate and we may be reluctant to give up on instinctive, if unformed, ideas of chance in all kinds of circumstances, whether repeatable or not. But as W. Feller has pointed out, "We may fairly lament that intuitive probability is insufficient for scientific purposes but it is a historical fact . . . . The appropriate, or 'natural,' probability distribution [for particles in statistical physics] seemed perfectly clear to everyone and has been accepted without hesitation by physicists. It turned out, however, that physical particles are not trained in human common sense and the 'natural' (or Boltzmann) distribution has to be given up for the ['unnatural' or 'non-intuitive'] Bose–Einstein distribution in some cases, for the Fermi–Dirac distribution in others."[3] The worth of an axiomatic model theory in mathematics is in the rich, unexpected theoretical developments and theorems that flow out of it; and its ultimate worth in application is its observed fit to empirical data and the correctness of its predictions. In these the modern theory of probability has been wildly successful—however unsettling some of its predictions to untrained intuition.

To illustrate the key features of the model it is best to begin with simple chance-driven situations with which we are readily familiar.

2 Chance experiments

Our intuitive assignment of probabilities to results of chance experiments is based on an implicit mathematical idealisation of the notion of repeated independent trials. For instance, in a coin-tossing experiment, conditioned by a complex of experience and custom, we are inclined to treat the coin as "fair"

[3] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume 1, 3rd Edition*, p. 5. © John Wiley & Sons, 1968. This material is reproduced with permission of John Wiley & Sons, Inc.

and to ascribe probabilities of 1/2 apiece for heads and tails ignoring possibilities such as that of the coin landing on edge or never coming down at all. Implicit here is the feeling that in a run of n tosses all 2^n possible sequences of heads and tails are equally likely to occur. If in a long run of n tosses there are m heads, we expect that the relative frequency m/n of the occurrence of heads in the tosses will be very close to 1/2, the accuracy getting better the larger n is.

Now, to be sure, no coin is really "fair". Statistical analyses of coin flips show invariably that heads and tails are *not* equally likely though the difference tends to be minute in most cases. Nonetheless, the mathematical fiction that a coin is fair is convenient in that it focuses on the essential features of the problem: it is not only simpler analytically but, for most applications, gives predictions that are sufficiently close to reality. We make similar assumptions about the throws of dice in the game of craps, the spin of a roulette wheel, or the distribution of bridge hands in cards. The following simple examples illustrate the key features of the modelling approach.

EXAMPLES: 1) *A coin is tossed three times.* Representing heads by $\mathfrak{H}$ and tails by $\mathfrak{T}$, the possible outcomes of the experiment may be tabulated in a natural convention as $\mathfrak{HHH}$, $\mathfrak{HHT}$, $\mathfrak{HTH}$, $\mathfrak{HTT}$, $\mathfrak{THH}$, $\mathfrak{THT}$, $\mathfrak{TTH}$, and $\mathfrak{TTT}$. It is clear that these are the only possible outcomes of the idealised experiment and, abeyant any reason to think otherwise, we suppose that all outcomes have equal chance 1/8 of occurrence. The event that exactly one head is seen may be identified with the aggregate of outcomes consisting of the sequences $\mathfrak{HTT}$, $\mathfrak{THT}$, and $\mathfrak{TTH}$ and it is natural to assign to this event the probability 3/8.

2) *The first throw in craps.* A classical die consists of six faces which we may distinguish by inscribing the numbers 1 through 6 on them (or, as is more usual, by inscribing one through six dots on the faces). The dice game of craps begins by throwing two dice and summing their face values. If the sum of face values is equal to 2, 3, or 12, the player loses immediately; if the sum is 7 or 11, she wins immediately; otherwise, the game continues. What are the chances that a player at craps loses on the first throw? wins on the first throw?

As the only element that decides the result of the first throw is the sum of face values it is natural to consider the outcomes of the experiment (as far as the first throw is concerned) to be the numbers 2 through 12. What are the chances we should ascribe to them? After a little thought the reader may come up with the numbers listed in Table 1. As a loss on the first throw is associated

Outcomes	2	3	4	5	6	7	8	9	10	11	12
Probabilities	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Table 1: *The sum of the face values of two dice.*

Cambridge University Press
9781107024472 - The Theory of Probability
Santosh S. Venkatesh
Excerpt
More information

with the aggregate $\{2, 3, 12\}$, it is now reasonable to ascribe to it the probability $\frac{1}{36} + \frac{2}{36} + \frac{1}{36} = \frac{1}{9}$. Similarly, a win on the first throw has associated with it the aggregate of outcomes $\{7, 11\}$ and accordingly has probability $\frac{6}{36} + \frac{2}{36} = \frac{2}{9}$. As any craps player knows, it is twice as likely that she wins on the first throw as that she loses on the first throw.

The critical reader may question the model for the experiment and may prefer instead a model of outcomes as ordered pairs of values, one for each die, the outcomes now ranging over the 36 equally likely possibilities, $(1, 1)$, $(1, 2)$, $\dots$, $(6, 6)$. In this model space, the event of a loss on the first throw may be associated with the aggregate consisting of the pairs $(1, 1)$, $(1, 2)$, $(2, 1)$, and $(6, 6)$, that of a win on the first throw with the aggregate of pairs $(1, 6)$, $(2, 5)$, $(3, 4)$, $(4, 3)$, $(5, 2)$, $(6, 1)$, $(5, 6)$, and $(6, 5)$. The corresponding probabilities then work out again to be $1/9$ and $2/9$, respectively. A variety of models may describe an underlying chance experiment but, provided they all capture the salient features, they will make the same predictions. All roads lead to Rome. ▶

The language of coins, dice, cards, and so on is picturesque and lends colour to the story. But in most cases these problems can be reduced to that of a prosaic placement of balls in urns. The following simple illustration is typical.

EXAMPLE 3) *An urn problem.* Two balls, say $\mathfrak{a}$ and $\mathfrak{b}$, are distributed in three urns labelled, say, 1, 2, and 3. With the order of occupancy in a given urn irrelevant, the outcomes of the experiment are nine in number, assumed to be equally likely of occurrence, and may be tabulated in the form

$$\begin{gathered} \mathfrak{ab}|-|-, \quad -|\mathfrak{ab}|-, \quad -|-|\mathfrak{ab}, \\ \mathfrak{a}|\mathfrak{b}|-, \quad \mathfrak{a}|-|\mathfrak{b}, \quad \mathfrak{b}|\mathfrak{a}|-, \quad \mathfrak{b}|-|\mathfrak{a}, \quad -|\mathfrak{a}|\mathfrak{b}, \quad -|\mathfrak{b}|\mathfrak{a}. \end{gathered} \tag{2.1}$$

The event that the second urn is occupied is described by the aggregate of outcomes $\{-|\mathfrak{ab}|-, \mathfrak{a}|\mathfrak{b}|-, \mathfrak{b}|\mathfrak{a}|-, -|\mathfrak{a}|\mathfrak{b}, -|\mathfrak{b}|\mathfrak{a}\}$ and hence has probability $5/9$. ▶

The reader should be able to readily see how the coin and dice problems may be embedded into generic urn problems concerning the placement of n balls into r urns. She may find some fun and profit in figuring out an appropriate urn model for the following catalogue of settings: birthdays, accidents, target shooting, professions (or gender or age), occurrence of mutations, gene distributions, and misprints.

The chance experiments we have considered hitherto deal with a finite number of possible outcomes. But our experience equips us to consider situations with an unbounded number of outcomes as well.

EXAMPLE 4) *A coin is tossed until two heads or two tails occur in succession.* The outcomes may be tabulated systematically in order of the number of tosses before the experiment terminates, leading to the denumerably infinite list of outcomes in Table 2. If the reader does not immediately believe that the assignment of

Outcomes	$\mathfrak{H}\mathfrak{H}$	$\mathfrak{T}\mathfrak{T}$	$\mathfrak{T}\mathfrak{H}\mathfrak{H}$	$\mathfrak{H}\mathfrak{T}\mathfrak{T}$	$\mathfrak{H}\mathfrak{T}\mathfrak{H}\mathfrak{H}$	$\mathfrak{T}\mathfrak{H}\mathfrak{T}\mathfrak{T}$	$\mathfrak{T}\mathfrak{H}\mathfrak{T}\mathfrak{H}\mathfrak{H}$	$\mathfrak{H}\mathfrak{T}\mathfrak{H}\mathfrak{T}\mathfrak{T}$	...
Probabilities	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$	...

Table 2: *A sequence of coin tosses terminated when two successive tosses coincide.*

probabilities is reasonable, a heuristic justification may be provided by the argument that, if we consider a very long, finite sequence of tosses of length n, a fraction $1/4$ of all such sequences begin with $\mathfrak{H}\mathfrak{H}$ and likewise also with $\mathfrak{T}\mathfrak{T}$, a fraction $1/8$ of all such sequences begin with $\mathfrak{T}\mathfrak{H}\mathfrak{H}$ and also with $\mathfrak{H}\mathfrak{T}\mathfrak{T}$, and so on. Allowing n to go to infinity permits the consideration of any terminating sequence of heads and tails. The objection that the experiment cannot in practice last an infinite amount of time so that arbitrarily long sequences are unrealistic in the model may be met with some force by the observation that, for the given probabilities, the chances of requiring more than 100 tosses, say, before termination are $2 \cdot 2^{-100}$. As this will require about 10^{38} performances of the experiment before one such occurrence is detected, one could argue with some justification that most of the assigned probabilities have not been fairly tested. In any case, the reader may well feel that it is even more artificial to fix a stopping point *a priori*, say at 50 tosses, numerical probabilities so chosen as to simply forbid longer sequences by fiat. The practical justification of the model lies in the fact that the assigned probabilities gel nicely with data for sequences of length up to ten or so which carry most of the likelihood; and presumably also for longer sequences though experimental data are abeyant given the fantastically small chances of occurrence.

In this setting, the event that at least four tosses are required before the experiment terminates is captured by the denumerably infinite aggregate of outcomes $\mathfrak{H}\mathfrak{T}\mathfrak{H}\mathfrak{H}$, $\mathfrak{T}\mathfrak{H}\mathfrak{T}\mathfrak{T}$, $\mathfrak{T}\mathfrak{H}\mathfrak{T}\mathfrak{H}\mathfrak{H}$, and so on. The probability hence that at least four tosses are required is given by

$$\frac{2}{16} + \frac{2}{32} + \frac{2}{64} + \cdots = \frac{2}{16}\left(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots\right) = \frac{2}{16}\bigg/\left(1 - \frac{1}{2}\right) = \frac{1}{4},$$

as we identify the infinite sum with a geometric series in powers of $1/2$. Likewise, to the event that the experiment concludes on an odd-numbered trial we may ascribe the probability

$$\frac{2}{8} + \frac{2}{32} + \frac{2}{128} + \cdots = \frac{2}{8}\left(1 + \frac{1}{4} + \frac{1}{16} + \cdots\right) = \frac{2}{8}\bigg/\left(1 - \frac{1}{4}\right) = \frac{1}{3},$$

as we now encounter a geometric series in powers of $1/4$. It is natural to extend the idea of summing over a finite number of outcome probabilities to a denumerably infinite sum when events are comprised of a countably infinite number of outcomes. ▸

What are the main features we can discern from simple chance experiments of this form? We begin with a model for a *gedanken* experiment whose performance, perhaps only in principle, results in an idealised outcome from a family of possible outcomes. The first element of the model is the specification of an abstract *sample space* representing the collection of idealised outcomes of the thought experiment. Next comes the identification of a family of *events* of interest, each event represented by an aggregate of elements of the sample space. The final element of the model is the specification of a consistent scheme of assignation of *probabilities* to events. We consider these elements in turn.

3 The sample space

R. von Mises introduced the idea of a sample space in 1931[4] and while his frequency-based ideas of probability did not gain traction—and were soon to be overtaken by Kolmogorov's axiomatisation—the identification of the abstract sample space of a model experiment paved the way for the modern theory.

We shall denote by the uppercase Greek letter Ω an abstract sample space. It represents for us the collection of idealised outcomes of a, perhaps conceptual, chance experiment. The elements ω of Ω will be called *sample points*, each sample point ω identified with an idealised outcome of the underlying *gedanken* experiment. The sample points are the primitives or undefined notions of the abstract setting. They play the same rôle in probability as the abstract concepts of points and lines do in geometry.

The simplest setting for probability experiments arises when the possible outcomes can be enumerated, that is to say, the outcomes are either finite in number or denumerably infinite. In such cases the sample space is said to be *discrete*. The examples of the previous section all deal with discrete spaces.

Examples: 1) *A coin toss.* The simplest non-trivial chance experiment. The sample space consists of two sample points that we may denote $\mathfrak{H}$ and $\mathfrak{T}$.

2) *Three tosses of a coin.* The sample space corresponding to the experiment of Example 2.1 may be represented by the aggregate $\mathfrak{HHH}$, $\mathfrak{HHT}$, …, $\mathfrak{TTT}$ of eight sample points.

3) *A throw of a pair of dice.* The sample space consists of the pairs $(1, 1)$, $(1, 2)$, …, $(6, 6)$ and has 36 sample points. Alternatively, for the purposes of Example 2.2 we may work with the sample space of 11 elements comprised of the numbers 2 through 12.

4) *Hands at poker, bridge.* A standard pack of cards contains 52 cards in four *suits* (called spades, hearts, diamonds, and clubs), each suit containing 13 distinct

[4] A translation of his treatise was published in 1964: R. von Mises, *Mathematical Theory of Probability and Statistics*, Academic Press, New York.

cards labelled 2 through 10, jack, queen, king, and ace, ordered in increasing *rank* from low to high. In bridge an ace is high card in a suit; in poker an ace counts either as high (after king) or as low (before 2). A poker hand is a selection of five cards at random from the pack, the sample space consisting of all $\binom{52}{5}$ ways of accomplishing this. A hand at bridge consists of the distribution of the 52 cards to four players, 13 cards per player. From a formal point of view a bridge hand is obtained by randomly partitioning a 52-card pack into four equal groups; the sample space of bridge hands hence consists of $(52)!/(13!)^4$ sample points. In both poker and bridge, the number of hands is so large that repetitions are highly unlikely; the fresh challenge that each game presents contributes no doubt in part to the enduring popularity of these games.

5) *The placement of two balls in three urns.* The sample space corresponding to Example 2.3 may be represented by the aggregate of points (2.1).

6) *The selection of a random graph on three vertices.* A graph on three vertices may be represented visually by three points (or *vertices*) on the plane potentially connected pairwise by lines (or *edges*). There are eight distinct graphs on three vertices—one graph with no edges, three graphs with one edge, three graphs with two edges, and one graph with three edges—each of these graphs constitutes a distinct sample point. A random graph (traditionally represented G_3 instead of ω in this context) is the outcome of a chance experiment which selects one of the eight possible graphs at random. Random graphs are used to model networks in a variety of areas such as telecommunications, transportation, computation, and epidemiology.

7) *The toss of a coin until two successive outcomes are the same.* The sample space is denumerably infinite and is tabulated in Example 2.4. Experiments of this stripe provide natural models for waiting times for phenomena such as the arrival of a customer, the emission of a particle, or an uptick in a stock portfolio. ▶

While probabilistic flavour is enhanced by the nature of the application at hand, coins, dice, graphs, cards, and so on, the theory of chance itself is independent of semantics and the specific meaning we attach in a given application to a particular outcome. Thus, for instance, from the formal point of view we could just as well view heads and tails in a coin toss as 1 and 0, respectively, without in any material way affecting the probabilistic statements that result. We may choose hence to focus on the abstract setting of discrete experiments by simply enumerating sample points in any of the standard ways (though tradition compels us to use the standard notation for these spaces instead of Ω).

EXAMPLES: 8) *The natural numbers* $\mathbb{N}$. The basic denumerably infinite sample space consists of the natural numbers $1, 2, 3, \dots$.

9) *The integers* $\mathbb{Z}$. Another denumerably infinite sample space consisting of integer-valued sample points $0, \pm1, \pm2, \dots$.