DATE 13 04 2020

1. The Sample Space

1. The Empirical Background.

The mathematical theory of probability gains practical value and an intuitive meaning in connection with real or conceptual experiments such as tossing a coin once, tossing a coin 100 times, throwing three dice, arranging a deck of cards, matching two decks of cards, playing Roulette, observing the lifespan of a radioactive atom or a person, selecting a random sample of people and observing the number of left-handers in it, crossing two species of plants and observing the phenotypes of the offspring; or with phenomena such as the sex of a newborn baby, the number of busy trunklines in a telephone exchange, the number of calls on a telephone, random noise in an electrical communication system, routine quality control of a production process, frequency of accidents, the position of a particle under diffusion. All these descriptions are rather vague and in order to render the theory meaningful, we have to agree on what we mean by possible results of the experiment or observation in question.

When a coin is tossed, it does not necessarily fall to heads or tails; it can roll away or stand on its edge. Nevertheless, we shall agree to regard "head" and "tail" as the only possible outcomes of the experiment. This convention simplifies the theory without affecting its applicability. Idealizations of this type are standard practice. It is impossible to measure the lifespan of an atom or a person without some error, but for theoretical purposes it is expedient to imagine that these quantities are exact numbers. The question then arises as to which numbers can actually represent the lifespan of a person. Is there a maximal age beyond which life is impossible, or is any age conceivable? We hesitate to admit that man can grow 1000 years old, and yet current

actuarial practice admits no bounds to the possible duration of life. According to the formulas on which modern mortality tables are based, the proportion of men surviving 1000 years is of the order of magnitude one in $10^{10^{36}}$. — a number with 10^{27} billions of zeros. This statement does not make sense from a biological or ~~statis~~ sociological standpoint, but considered exclusively from a statistical standpoint it certainly does not contradict any experience. There are fewer than 10^{10} people born in a century. To test the contention statistically, more than $10^{10^{35}}$ centuries would be required, which is considerably more than $10^{10^{34}}$ lifetimes of the earth. Obviously, such extremely small probabilities are compatible with our notion of impossibility. Their use may appear utterly absurd; but it does no harm and is convenient in simplifying many formulas. Moreover if we were to seriously discard the possibility of living 1000 years, we should have to accept the existence of a maximum age, and the assumption that it should be possible to live n years and impossible to live n years and two seconds is as unappealing as the idea of unlimited life.

Any theory necessarily involves idealizations, and our first idealization concerns the possible outcomes of an "experiment" or "observation". If we want to construct an abstract model, we must at the outset reach a decision about what constitutes a possible outcome of the (idealized) experiment.

For uniform terminology, the results of experiments

DATE

or observations will be called events. Thus, we shall speak of the event that of the five coins tossed more than three fell heads. similarly, the experiment of distributing the cards in bridge may result in the event that North has two aces. The composition of a sample ("two left-handers in a sample of 85") and the result of a measurement ("temperature 120°", "seven sunshine days") each will be called an event.

We shall distinguish between compound (decomposable) and simple (or indecomposable) events. For example, saying that a throw with two dice result in "sum six" amounts to saying that it resulted in "(1,5) or (2,4) or (3,3) or (4,2) or (5,1)" and this enumeration decomposes the event "sum six" into five simple events. similarly, the event "two odd faces" admits of the decomposition "(1,1) or (1,3) or ... or (5,5)" into nine simple events. Note that, if a throw results in (3,3), then the same throw results also in the events "sum six" and "two odd faces"; these events are not mutually exclusive and hence may occur simultaneously. As a second example, consider the age of a person. Every particular value x represents a simple event, whereas the statement that a person is in his fifties describes the compound event that x lies between 50 and sixty. In this way, every compound event can be decomposed into simple events and vice versa, that is to say, a compound event is an aggregate of certain simple events.

If we want to speak about experiments or observations in a theoretical way and without ambiguity, we must

DATE

first agree on simple events representing the thinkable outcomes; they define the idealized experiment. It is usual to refer to these simple events as sample points, or points for short. By definition, every indecomposable result of the (idealized) experiment is represented by, and only one sample point. The aggregate of all sample points will be called sample space. All events connected with a given (idealized) experiment can be described in terms of sample points.

Before formalizing these basic conventions, we proceed to discuss a few typical examples which will play a role further on.

2. Examples.

(a) Distribution of three balls in three cells.

Table 1 describes all possible outcomes of the experiment of placing three balls into three cells.

Table 1.

1. {abc | - | - }	10. {ac | b | - }	19. {b | ac | - }
2. { - | abc | - }	11. {ac | - | b }	20. {b | - | ac }
3. { - | - | abc }	12. { - | ac | b }	21. { - | b | ac }
4. {ab | c | - }	13. { c | ab | - }	22. { a | b | c }
5. {ab | - | c }	14. { c | - | ab }	23. { b | a | c }
6. { - | ab | c }	15. { - | c | ab }	24. { a | c | b }
7. {bc | a | - }	16. { a | bc | - }	25. { c | b | a }
8. {bc | - | a }	17. { a | - | bc }	26. { b | c | a }
9. {bc | bc | a }	18. { - | a | bc }	27. { c | a | b }.

Each of these arrangements represents a simple event, that is, a sample point. The event A, "one cell is multiply occupied"

is realized in arrangements 1-21, and we express this by saying that the event A is the aggregate of the sample points 1-21. Similarly, the event B "first cell is not empty" is the aggregate of the sample points 1, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22-27. The event C defined by "both A and B occur" is the aggregate of the thirteen sample points 1, 4, 5, 7, 8, 10, 11, 13 14, 16, 17, 19 and 20. In this particular example it so happens that each of the 27 points belongs to either A or B (or to both); therefore the event "either A or B occurs" is the entire sample space and occurs with absolute certainty. The event D defined by A does not occur consists of the points 22-27 and can be described by the condition that no cell remains empty. The event "first cell empty and no cell multiply occupied" is impossible (does not occur) since no sample point satisfies these specifications.

(b) Distribution of r balls in n cells. The more general case of r balls in n cells can be studied in the same manner, except that the number of possible arrangements increases rapidly with r and n. For $r = 3$ balls and $n = 4$ cells, the sample space already contains $4^3 = 64$ points, and for $r = n = 10$, there are 10^{10} sample points; a complete tabulation would require some hundred thousand big volumes.

We use this example to illustrate the important fact that the nature of sample points is irrelevant for our theory. To us, the sample space (together with the probability distribution defined on it) defines the idealized experiment. We use the picturesque language of balls and cells, but the same sample space admits of a great variety of different practical interpretations. To clarify this point, and also for further reference, we list

here a number of situations in which the intuitive background varies; all are, however, abstractly equivalent to the scheme of placing r balls into n-cells, in the sense that the outcomes differ only in their verbal descriptions. The appropriate assignment of probabilities is not the same in all cases and will be discussed later on.

(b, 1). Birthdays. The possible configurations of the birthdays of r people correspond to the different arrangements of r balls in n=365 cells (assuming the year to have 365 days).

(b, 2) Accidents. Classifying r accidents according to the weekdays when they occurred is equivalent to placing r balls into n=7 cells.

(b, 3) r× firing at n targets, the hits correspond to balls, the targets to cells.

(b, 4) Sampling. Let a group of r people be classified according to say, age or profession. The classes play the role of cells, and people those of balls.

(b, 5) Irradiation in biology. When the cells in the retina of the eye are exposed to light, the light particles play the role of the balls and the actual cells are the cells of our model. Similarly, in studying of the body of the genetic effects of irradiation, the chromosomes correspond to the cells of our model and α-particles to the balls.

classmate

(6,6) In cosmic ray experiments the particles hitting the Geiger counters represent the balls and the counters function as cells.

(6,7) An elevator starts with r passengers and stops at n floors. The different arrangements of discharging the passengers are replicas of the different distributions of r balls in n cells.

(6,8) Dice. The possible outcomes of a throw of r dice correspond to placing r balls into $n=6$ cells. When tossing a coin we are in effect dealing with only $n=2$ cells.

(6,9) Random digits. The possible orderings of a sequence of r digits correspond to the distribution of r balls (= places) into ten cells called $0, 1, 2, \ldots, 9$.

(6,10) The sex distribution of r persons. Here we have $n=2$ cells and r balls.

(6,11) Coupon collecting. The different kinds of coupons represent the cells; the coupons collected represent the balls.

(6,12). Deals in bridge. The four players represent 4 cells and we have $r=4$ balls.

(6,13) Gene distributions. Each descendant of an individual (person, plant or animal) inherits from the progenitor certain genes. If a particular gene can appear in ~~is~~ n forms $A_1, \ldots, A_n$ then the descendants may be classified according

DATE

to the type of gene. The descendants correspond to the balls, the genotypes $A_1, \ldots, A_{20}$ to the cells.

(b,14) Chemistry. Suppose a long chain polymer reacts with oxygen. An individual chain may react with $0, 1, 2, \ldots,$ oxygen molecules. Here, the reacting oxygen molecules play the role of balls and the polymer chains play the role of cells into which balls are put.

(b, 15). Theory of photographic emulsions. A photographic plate is covered with grains sensitive to light quanta: a grain reacts if it is hit by a certain number n of the quanta. For the theory of black-white contrast we must know, how many cells are likely to be hit by the n quanta. We have here an occupancy problem where the grains correspond to the cells and the light quanta to the balls. (Actually the situation is much more complicated since a plate usually contains grains of different sensitivity).

(b,16) Misprints. The possible distributions of n misprints in the r pages of a book correspond to all the different distributions of n balls in r cells, provided n is smaller than the number of letters per page.

(c) The case of indistinguishable balls. Let us return to example (a) and suppose that the three balls are not distinguishable. This means that we no longer distinguish between those arrangements such as 4, 7, 10. The table thus reduced to the below:

classmate

PAGE

DATE

1. {***|-|-}
2. {**|*|-}
3. {**|-|*}
4. {*|**|-}
5. {*|-|**}
6. {-|*|**}
7. {-|***|-}
8. {-|-|***}
9. {-|**|*}
10. {*|*|*}.

The latter defines the sample space of the ideal experiment which we call placing three indistinguishable balls into three cells", and a similar procedure applies to the case of n balls in r cells.

Whether or not actual balls in practice are indistinguishable is irrelevant for our theory. Even if they are, we may decide to treat them as indistinguishable.

Note. A deck of bridge cards consists of 52 cards, arranged in four suits of thirteen each. There are thirteen face values (2, 3, ..., 10, Jack, Queen, King, Ace) in each suit. The four suits are called spades, clubs, hearts and diamonds. The last two are red, the first two black. Cards of the same face value are called of the same kind. For our purposes, playing bridge means distributing the cards to four players, to be called North, South, East and West (or N, S, E, W for short) so that each receives thirteen cards. Playing poker by definition, means selecting five cards out of the pack.

The ~~ones~~ in

The ones in bridge or the people in our elevator certainly are distinguishable and yet it is often preferable to treat them as indistinguishable. ~~The ones in bridge or the people in our elevator certainly are distinguishable, and yet it is often~~

PAGE

classmate

The dice of example (b,8) may be colored to make them distinguishable, but whether in discussing a particular problem we use the model of distinguishable or indistinguishable balls is purely a matter of purpose and convenience. The nature of a concrete problem may dictate the choice, but under any circumstances our theory begins only after the appropriate model has been chosen, that is, after the sample space has been defined.

In the scheme above, we have considered indistinguishable balls, but the table 2 still refers to a first, second, third cell and their order is essential. We can go a step further and assume that even the cells are indistinguishable (for example a cell may be chosen at random without regard to its contents). With both balls and cells indistinguishable (for example only three different arrangements are possible, namely $\{***|-|-\}$, $\{**|*|-\}$, $\{*|*|*\}$.

(d) Sampling. Suppose that a sample of 100 people is taken in order to estimate how many people smoke. The only property of the sample of interest in this connection is the number x of smokers; this may be any integer between 0 and 100. In this case we may agree that our sample space consists of the 101 points $0, 1, \ldots, 100$. Every particular sample or observation is completely described by stating the corresponding point x. An example of a compound event is the result that the majority of the people sampled are

smokers. This means that the experiment resulted one of the fifty simple events $50, 51, \ldots, 100$, but it is not stated in which. Similarly, every property of the sample can be described in enumerating the corresponding cases or sample points. For uniform terminology, we speak of events rather than properties of the sample. Mathematically, an event is simply the aggregate of the corresponding sample points.

(e) Sampling (continued).
Suppose now that 100 people in our sample are classified not only as smokers or non-smokers but also as males or females. The sample may now be characterised by a quadruple (M_s, F_s, M_n, F_n) of integers giving in order the number of male and female smokers, male and female non-smokers. We can take for sample points the quadruples of integers lying between 0 and 100 and adding to 100. There are 176,851 such quadruples and they constitute the sample space. The event relatively more males than [fewer] females smoke means that in our sample the ratio M_s/M_n is greater than F_s/F_n. The point (73, 2, 8, 17) has this property, but (0, 1, 50, 49) has not. Our event can be described by enumerating all quadruples with the desired property.

(f) Coin tossing. For the experiment of tossing a coin three times, the sample space consists of eight points which may conveniently be represented by HHH, HHT, HTH, HTT, THH, THT, TTH, TTT. The event A, "two or more heads" is the aggregate of the first four points. The event B, just one tail means either HHT, or HTH; we say that B contains these three points.

(g) Ages of a couple. An insurance company is interested in the age distribution of couples. Let x stand for the age of the husband, y for the age of the wife. Each observation results in a number-pair (x, y). For the sample space corresponding to a single observation, we take the first quadrant of the xy-plane, so that each point $x>0, y>0$ is a sample point. The event A, "husband is older than 40", is represented by all points to the right of the line $x=40$; the event B, "husband is older than wife, is represented by the angular region between the x-axis and the bisector $y=x$, that is to say, by the aggregate of points with $x>y$; the event C, "wife is older than 40", is represented by the portion of the first quadrant above the line $y=40$. For a geometric representation of the joint age distributions of two couples we would require a four-dimensional space.

(h) Phase space. In statistical mechanics, each possible state of a system is called a point in "phase space." This is only a difference in terminology. The phase space is simply our sample space; its points are our sample points.

3. The sample space - events.

It should be clear from the preceding that we shall never speak of probabilities except in relation to a given sample space (or physically in relation to a given conceptual experiment). We start with the notion of a sample space and its points; from now on they will be considered given. They are primitive