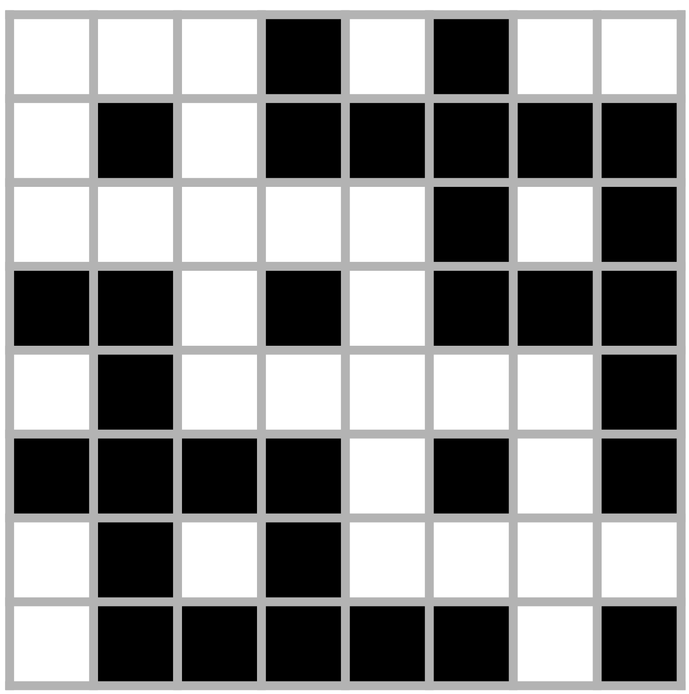
ELEMENTARY PROBABILITY

Thomas Wieting

Reed College

2015



ELEMENTARY PROBABILITY

1 Introduction

01° The object of this booklet is to introduce the subject of probability theory by means of examples. In the beginning, the examples and the relevant computations will be quite easy; in the end, quite hard. The examples involve nothing more than finite sets and the computations require nothing more than basic arithmetic. Hence, to understand these matters, one need have no special preparation. But one must bring forward that cheerfully implacable attitude of mind which, in general, characterizes the serious study of Mathematics.

02° There are two special features of this booklet which we hope will stimulate interest. The first involves a paradox in the interpretation of (conditioned) probability, described in Section 6. The second involves an application to the "real world" of weaving, described in Section 7.

03° Note that the articles are numbered consecutively, for ready reference. The articles labelled j° compose the text; the articles labelled $k^\bullet$ compose the problem set.

04° We all know how to use such sentences as the following, in ordinary conversation:

(•) She is more likely than he to solve the problem.
(•) I am virtually certain to attend Oregon State University.
(•) The odds on rolling a seven with two fair dice are 1 to 5.
(•) My chances of winning the lottery are nil.

In each case (though in different ways), we apply the intuitive idea of *probability*. However, it is very difficult to put forward a sharp definition of this intuitively familiar term. Fortunately, we may study and apply probability theory without such a definition. In the fashion of modern Mathematics, we invoke the basic components of set theory to build a *model* of the idea of probability about which we can make precise statements and prove useful theorems. Of course, one is free to argue that a given model is inadequate and one is free to propose another. Models are judged by the criteria of elegance and utility.

We plan to describe a very simple model of the idea of probability, which involves just finitely many possible *events*.

05° With due respect, let us note that the first general, precise model of the idea of probability was put forward by the celebrated Russian mathematician A. Kolmogoroff. His work was translated into German:

Grundbegriffe der Wahrsheinlichkeitsrechnung

and communicated to the Western World in 1933. We plan to describe the simpler aspects of Kolmogoroff's model.

06° Let X be a finite set. We will refer to the various subsets A of X as *events*. To each event A, let us assign a real number $P(A)$. We will refer to $P(A)$, at this point rather cryptically, as the *probability* of A. We require that the following conditions be met:

(•) for each event A, $0 \le P(A) \le 1$
(•) $P(\emptyset) = 0$ and $P(X) = 1$
(•) for any events A' and A'', if $A' \cap A'' = \emptyset$ then $P(A' \cup A'') = P(A') + P(A'')$

We will refer to the foregoing assembly, composed notably of the finite set X and the probability function P defined on the subsets of X, as a finite *probability space*.

07$^\bullet$ Let the number of members of X be n. Explain why the number of events must be 2^n.

08° Let x be any member of X. In practice, one identifies x with the event $\{x\}$ composed of the single member x. One refers to x as an *elementary event*. By the required conditions for P, it is plain that, for each event A, $P(A)$ is the sum of the probabilities corresponding to the elementary events which are contained in A:

$$P(A) = \sum_{x \in A} P(x)$$

Therefore, we may describe a finite probability space simply by displaying the elementary events and the corresponding probabilities, in something like the following manner:

$$x_1, x_2, x_3, \ \dots\ , x_n$$

$$p_1, p_2, p_3, \ \dots\ , p_n$$

where:

$$p_j := P(x_j) \qquad (1 \leq j \leq n)$$

That will be our usual practice.

09° Of course, we require flexible notation to make interesting displays. For example, let us describe the *bernoulli trial* space, a finite probability space for which the elementary events are:

$$0,\ 1$$

and the corresponding probabilities are:

$$p_0,\ p_1$$

where p_0 and p_1 are any real numbers for which $0 \leq p_0 \leq 1$, $0 \leq p_1 \leq 1$, and $p_0 + p_1 = 1$. Why? One interprets 0 as *failure* and 1 as *success*.

10° Let us describe the *dice roll* space, an example which will figure in many subsequent articles. The elementary events are:

$$\begin{matrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{matrix}$$

and the corresponding probabilities are:

$$\begin{array}{cccccc}
p_{(1,1)} & p_{(1,2)} & p_{(1,3)} & p_{(1,4)} & p_{(1,5)} & p_{(1,6)} \\
p_{(2,1)} & p_{(2,2)} & p_{(2,3)} & p_{(2,4)} & p_{(2,5)} & p_{(2,6)} \\
p_{(3,1)} & p_{(3,2)} & p_{(3,3)} & p_{(3,4)} & p_{(3,5)} & p_{(3,6)} \\
p_{(4,1)} & p_{(4,2)} & p_{(4,3)} & p_{(4,4)} & p_{(4,5)} & p_{(4,6)} \\
p_{(5,1)} & p_{(5,2)} & p_{(5,3)} & p_{(5,4)} & p_{(5,5)} & p_{(5,6)} \\
p_{(6,1)} & p_{(6,2)} & p_{(6,3)} & p_{(6,4)} & p_{(6,5)} & p_{(6,6)}
\end{array}$$

More succinctly, the elementary events are:

$$(j,k) \qquad (1 \leq j \leq 6, 1 \leq k \leq 6)$$

and the corresponding probabilities are:

$$p_{(j,k)} \qquad (1 \leq j \leq 6, 1 \leq k \leq 6)$$

We require that:

$$0 \leq p_{(j,k)} \leq 1 \qquad (1 \leq j \leq 6, 1 \leq k \leq 6)$$

and that:

$$\sum_{j=1}^{6}\sum_{k=1}^{6} p_{(j,k)} = 1$$

Why? For any indices j $(1 \leq j \leq 6)$ and k $(1 \leq k \leq 6)$, one interprets (j,k) as the outcome following the roll of two dice in which the first die comes up j and the second die comes up k. For fair dice, one would expect:

$$p_{(j,k)} = 1/36 \qquad (1 \leq j \leq 6, 1 \leq k \leq 6)$$

Why?

11° Let us describe the *magnetic string* space, based upon three atoms:

$$\bullet \; \bullet \; \bullet$$

each of which can be oriented (so to speak) downward or upward. For this finite probability space, the elementary events are:

$$\downarrow\downarrow\downarrow \;\; \downarrow\downarrow\uparrow \;\; \downarrow\uparrow\downarrow \;\; \downarrow\uparrow\uparrow \;\; \uparrow\downarrow\downarrow \;\; \uparrow\downarrow\uparrow \;\; \uparrow\uparrow\downarrow \;\; \uparrow\uparrow\uparrow$$

and the corresponding probabilities are:

$$p_{\downarrow\downarrow\downarrow} \;\; p_{\downarrow\downarrow\uparrow} \;\; p_{\downarrow\uparrow\downarrow} \;\; p_{\downarrow\uparrow\uparrow} \;\; p_{\uparrow\downarrow\downarrow} \;\; p_{\uparrow\downarrow\uparrow} \;\; p_{\uparrow\uparrow\downarrow} \;\; p_{\uparrow\uparrow\uparrow}$$

Of course, we impose the usual requirements on the probabilities.

With patience, one may develop the magnetic string space for (not necessarily three but for) any finite number of atoms. In realistic applications (to thermodynamics), the number is enormous and the arrangement is very complex.

12° Now let X be a finite probability space and let P be the associated probability function defined on the subsets of X. Let A be any event. That is, let A be any subset of X. Naturally, we expect to be able to say that A *occurs* (or that it does not) and that $P(A)$ is the *probability* that A occurs. The foregoing examples will help to interpret these expressions.

In the first case, we imagine that an Arbitrator (let him be Bernoulli) flips a coin, for which the outcome x is either 0 (tails, failure) or 1 (heads, success). In the second case, we imagine that a Croupier rolls two dice. The outcome $x = (j, k)$ indicates which of the faces of the two dice came up. In the last case, we imagine that Nature assembles the magnetic string: by a complex interplay of forces, the particular outcome x emerges.

In general, we imagine that Tyche, the goddess of Chance, reaches into the set X and draws out a member, let it be x, at random, then returns the member to the set right away.

In any case, we say that A “occurs” iff $x \in A$.

13° Now we say that $P(A)$ is the probability that an elementary event x, chosen at random in X, is contained in A. Responsibility for the choice of x lies with Tyche. We presume that if Tyche makes many such choices, one after the other “independently,” she will do so in such a way that:

$$P(A) \approx \frac{M}{N}$$

where N is the total number of her choices and where M is the number of her choices which lie in A.

14° One may clarify the foregoing rather obscure commentary by studying one of the fundamental theorems of Probability Theory, the Strong Law of Large Numbers. But that is a matter for another day.

15° Again let X be a finite probability space and let P be the associated probability function defined on the subsets of X. Let A be any event. Let:

$$x_1, x_2, x_3, \ \dots \ , x_n$$

be the elementary events and let:

$$p_1, p_2, p_3, \ \dots \ , p_n$$

be the corresponding probabilities. In many interesting cases, the various elementary events are *equally likely*, which is to say that:

$$p_1 = p_2 = p_3 = \cdots = p_n = 1/n$$

In such cases:

$$P(A) = \frac{m}{n}$$

where m is the number of members of A. One says that $P(A)$ is the ratio of the number m of "favorable" elementary events to the number n of all possible elementary events.

For finite probability spaces X in which the elementary events are equally likely, the computation of the probability of a (general) event A reduces to problems in counting, specifically, the problems of counting the numbers of members of X and A. However, such counting problems can be very complicated. We will devote the next section to a discussion of some of the basic principles of counting, which will prove useful in subsequent computations of probabilities.

16$^\bullet$ For the case of the dice roll, assume that the elementary events are equally likely and compute the probability that the Croupier rolls a seven.

17$^\bullet$ For the case of the magnetic string with three atoms, assume that the elementary events are equally likely. Assign to each elementary event x an integer H equal to the difference between the number of upward oriented atoms and the number of downward oriented atoms. Note that the values of H can be -3, -1, 1, and 3 but none other. Calculate the probabilities of occurence for each of the values of H.

18$^\bullet$ Recently, a monkey purchased a typewriter with 13 keys:

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, \times, =$$

He understood nothing of the meaning of the symbols but he soon found that he enjoyed typing. Repeatedly, he typed strings of symbols, such as:

$$6\,6\,1 = \quad \text{and} \quad 9 = 0\,0 \times \times$$

In time, he fell into the practice of typing only strings of length five. Presume that he continues the practice. Calculate the probability that, on any particular instance of typing, the monkey types a "true equation," such as:

$$4 + 5 = 9$$

19° Let X' and X'' be any finite sets. Let $X' \times X''$ be the product of the two sets X' and X''. The members of $X' \times X''$ are the ordered pairs:

$$(x', x'')$$

where x' is any member of X' and where x'' is any member of X''. Clearly, if the number of members of X' is n' and if the number of members of X'' is n'' then the number of members of $X' \times X''$ is $n' \times n''$. This statement is the Product Principle.

20• Explain how the Product Principle figures in the solution of article 18•.

21• Apply the Product Principle to solve the following Birthday Problem. Let m be any positive integer. Imagine selecting m people at random. Explain why the number:

$$Q_m := \frac{365 \times (365-1) \times \cdots \times (365-m+1)}{365^m}$$

equals the probability that no two of the selected people have the same birthday. Of course:

$$P_m := 1 - Q_m$$

equals the probability that some two of the selected people have the same birthday. Calculate:

$$P_2, P_3, P_4, \ \ldots$$

until you find the first value of m for which:

$$\frac{1}{2} < P_m$$

The point of the Birthday Problem is that the discovered value of m is surprisingly small.

The following display shows a Mathematica program for computing the required numbers, and a table of the first forty of the numbers.

```
jm=39;
BQ={1};
Do[{xx=Take[BQ,-1],yy=xx*(365-jj)/365;BB=Append[BQ,yy],
BQ=BB},
{jj,jm}];
BR=1-N[BQ];
BP=Table[{jj,BR[[jj]]},{jj,jm+1}];
TableForm[BP]
```

1	0.
2	0.00273973
3	0.00820417
4	0.0163559
5	0.0271356
6	0.0404625
7	0.0562357
8	0.0743353
9	0.0946238
10	0.116948
11	0.141141
12	0.167025
13	0.19441
14	0.223103
15	0.252901
16	0.283604
17	0.315008
18	0.346911
19	0.379119
20	0.411438
21	0.443688
22	0.475695
23	0.507297
24	0.538344
25	0.5687
26	0.598241
27	0.626859
28	0.654461
29	0.680969
30	0.706316
31	0.730455
32	0.753348
33	0.774972
34	0.795317
35	0.814383
36	0.832182
37	0.848734
38	0.864068
39	0.87822
40	0.891232

22° Let X be any finite set and let n be the number of members of X. For example, let $n = 3$ and let the members of X be the following:

$$u, v, w$$

In the foregoing display, we have listed the members of X in an apparent order: first, u; second, v; third, w. Just as well, we could have listed the members of X in various other apparent orders:

$$u, w, v \quad v, u, w \quad v, w, u \quad w, u, v \quad w, v, u$$

By the Product Principle, there are $3! := 3 \times 2 \times 1$ such ordered lists. Why? In the general case, the number of ordered lists of the members of X is:

$$n! := n \times (n-1) \times \ \cdots \ \times 2 \times 1$$

This statement is the Ordered List Principle.

$23^\bullet$ Compute 2!, 3!, 4!, 5!, and so forth, to develop a sense of the rapid growth of the *factorials* $n!$.

24° Let X be any finite set and let n be the number of members of X. Let j be an integer for which $0 \leq j \leq n$. Among the subsets Y of X, there are some which contain j members. We plan to determine the number of such j-member subsets.

To form a j-member subset Y of X, we first select any one of the n members of X. We then select any one of the remaining $n-1$ members of X. We continue until we have made a full run of j selections. By the Product Principle, the number of such runs is:

$$n \times (n-1) \times (n-2) \times \ \cdots \ \times (n-j+1)$$

Of course, each such run determines a j-member subset Y of X. However, two such runs may determine the same j-member subset Y of X because the runs determine the subsets as ordered lists. To obtain the number of j-member subsets Y of X, we must divide the number of runs by $j!$:

$$\binom{n}{j} := \frac{n \times (n-1) \times (n-2) \times \ \cdots \ \times (n-j+1)}{j \times (j-1) \times \ \cdots \ 2 \times 1} = \frac{n!}{j! \times (n-j)!}$$

Hence, the number of j-member subsets Y of X is the *binomial coefficient*:

$$\binom{n}{j} = \frac{n!}{j! \times (n-j)!}$$

This statement is the Selection Principle.

For ease of expression, one reads the foregoing binomial coefficient as "n choose j."

25° The notational conventions:

$$0! := 1 \quad \text{and} \quad 1! := 1$$

cover the slightly ambiguous cases in which j equals 0, 1, $n-1$, or n and lead to correct values for the number of j-member subsets Y of X.

26• For the case of the magnetic string with twelve atoms, assume that the elementary events are equally likely. Assign to each elementary event x an integer H equal to the difference between the number of upward oriented atoms and the number of downward oriented atoms. Note that the values of H can be -12, -10, ..., 10, and 12 but none other. Calculate the probability that H equals -10 or 4.

27° Finally, let us describe the Placement Principle. We will lapse now into picturesque terms, designed to suggest the broad range of possible applications for this principle.

Let j and k be any positive integers. We imagine k indistinguishable objects and j distinguishable boxes. To be explicit, we assume that the boxes are labelled in a specific order:

$$1, 2, 3, \ \ldots \ , j$$

We plan to count the number of ways by which one may place the objects in the boxes.

For example, let $j = 5$ and let $k = 12$. Consider the following string of zeros and ones:

$$1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1$$

We interpret the foregoing string as a complete description of a particular way by which one may place the 12 indistinguishable objects in the 5 distinguishable boxes, as follows. We shall place 3 objects in the first box, 4 objects in the second box, 1 object in the third box, 0 objects in the fourth box, and 4 objects in the fifth box. Why?

How shall the following strings be interpreted:

$$1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0$$

$$1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1$$

?

Clearly, the number of ways by which one may place 12 indistinguishable objects in 5 distinguishable boxes equals the number of such strings.

To count the number of such strings, we note first that each string has length $5 - 1 + 12$ and contains (in varying positions) $5 - 1$ zeros and 12 ones. Now let us consider the (finite) set X composed (for simplicity) of the following $5 - 1 + 12$ members:

$$a \quad b \quad c \quad d \quad e \quad f \quad g \quad h \quad i \quad j \quad k \quad l \quad m \quad n \quad o \quad p$$

We interpret the string:

$$1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1$$

as a complete description of a 12-subset Y of X, as follows. With regard to the array:

$$\begin{array}{cccccccccccccccc} a & b & c & d & e & f & g & h & i & j & k & l & m & n & o & p \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \end{array}$$

we shall take a member of X to be a member of Y iff that member lies above a one in the array. Hence, the (12) members of Y are the following:

$$a \quad b \quad c \quad e \quad f \quad g \quad h \quad j \quad m \quad n \quad o \quad p$$

Clearly, the number of such strings equals the number of 12-member subsets Y of X. By the Selection Principle, that number is:

$$\binom{5-1+12}{12}$$

Now, by generalization, we conclude that the number of ways by which one may place k indistinguishable objects in j distinguishable boxes equals:

$$\binom{j-1+k}{k} = \frac{(j-1+k)!}{(j-1)! \times k!}$$

This statement is the Placement Principle.

28• An old and successful gambler has 10 million dollars and 6 sons. In his will, he directs his lawyer to bequeath his millions to his sons in the following way. His lawyer shall construct a (large) roulette wheel with:

$$3003 = \binom{15}{10}$$

slots. The slots shall be marked with strings composed of 5 zeros and 10 ones so that the slots and strings stand in bijective correspondence. The lawyer shall set the roulette wheel in motion and, in the usual fashion, cast a marble into the wheel. He shall determine the slot in which the marble comes to rest. The corresponding string shall determine the assignment of the millions to the sons. How?

What is the probability that each son receives at least one million? What is the probability that the youngest (that is, the sixth) son receives half the fortune?

By the way, the gambler had initially instructed his lawyer to bequeath his millions to his sons in the following way. The lawyer shall roll ten fair dice. He shall assign to the first (that is, to the eldest) son, the number of millions equal to the number of ones which come up; to the second son, the number of millions equal to the number of twos which come up; and so forth. But his lawyer pointed out to him that the outcomes would not be equally likely. He confessed that he did not know how to calculate the probabilities for such a procedure and so could not assess whether or not the prodecure was "fair." Explain his reasoning.

To calculate probabilities under the latter procedure, one requires not just binomial but *multinomial* coefficients. We will discuss this matter informally in the class meetings.

4 Random Variables

29° Let X be a finite probability space and let P be the associated probability function defined on the subsets of X. Let Y be a finite set and let F be a *mapping* carrying X to Y. We mean to say that F is some sort of "rule" which assigns to each member x of X a corresponding member $F(x)$ of Y. By analogy with the familiar concept of *function* (for which the values are real numbers), we refer to $F(x)$ as the *value* of F at x.

We refer to F as a *random variable* defined on X with values in Y.

30° Now let us apply X, P, Y, and F to define a probability function Q on the subsets of Y. In this way, we obtain a new finite probability space Y.

The design of Q is very natural. Let B be any subset of Y. We imagine that Tyche draws out a member x from X at random. We inquire whether $F(x) \in B$. If so, we say that B "occurs." We wish to compute the probability that B occurs. To that end, let A be the subset of X composed of all members x for which $F(x) \in B$. By the foregoing observations, we are led to define $Q(B)$ to be $P(A)$.

One usually denotes the set A by $F^{-1}(B)$. With this notation, we may formally define the probability function Q as follows:

$$Q(B) := P(F^{-1}(B))$$

where B is any subset of Y. In the archaic terms of yesteryear, one refers to Q as the *distribution* of the random variable F.

31° In practice, we begin with a finite probability space X for which the elementary events are equally likely, introduce a random variable F, and obtain a new finite probability space Y. Many interesting finite probability spaces arise in this manner.

$32^\bullet$ With reference to article $26^\bullet$, show that one may interpret H as a random variable defined on the magnetic string space (with twelve atoms). The set Y would be composed of the following elementary events:

$$-12,\ -10,\ -8,\ -6,\ -4,\ -2,\ 0,\ 2,\ 4,\ 6,\ 8,\ 10,\ 12$$

For problem $26^\bullet$, $Q(-10)$ and $Q(4)$ were computed. Finish the problem by computing the complete probability function Q, that is, the distribution of H.

$33^\bullet$ Let X be the dice roll space with equally likely elementary events:

$$(j,k) \qquad (1 \le j \le 6, 1 \le k \le 6)$$

Let Y be the set composed of the following integers:

$$2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ 12$$

and let F be the random variable defined on X as follows:

$$F(j,k) := j + k \qquad (1 \le j \le 6, 1 \le k \le 6)$$

Calculate the distribution Q of F. By article $16^\bullet$, we already know that $Q(7) = 1/6$.

34° Now let us describe a particular family of random variables for which the corresponding distributions are the celebrated *binomial distributions*.

We begin with the bernoulli trial space X, for which the elementary events are:

$$0,\ 1$$

and the corresponding probabilities are:

$$p_0,\ p_1$$

Let n be any positive integer. Let $\bar{X}$ be the (finite) set composed of all finite strings of the form:

$$\bar{x} := \epsilon_1\epsilon_2\epsilon_3\ \ldots\ \epsilon_n$$

where, for each index j $(1 \leq j \leq n)$, ϵ_j equals either 0 or 1. Of course, the number of members $\bar{x}$ of $\bar{X}$ is $\bar{n} := 2^n$. We interpret such a string $\bar{x}$ as a run of trials. For each index j $(1 \leq j \leq n)$, ϵ_j if 0 signifies failure and if 1 signifies success. We define the corresponding probability as follows:

$$p_{\bar{x}} := p_{\epsilon_1} \times p_{\epsilon_2} \times p_{\epsilon_3} \times \cdots \times p_{\epsilon_n}$$

From these real numbers, we can assemble the probability function $\bar{P}$ defined on the subsets of $\bar{X}$. [For a review, see article 8°.] The resulting probability space $\bar{X}$ is the *bernoulli n-trial* space.

35° Let $n = 16$. Let $\bar{x}$ be the following string:

$$\bar{x} := 1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1$$

Note that:

$$p_{\bar{x}} = p_0^4 \times p_1^{12}$$

36° Now let $\bar{Y}$ be the following (finite) set of nonnegative integers:

$$0, 1, 2, \ldots, n$$

Let $\bar{F}$ be the mapping carrying $\bar{X}$ to $\bar{Y}$ which assigns to each $\bar{x}$ in $\bar{X}$ the number of 1s in $\bar{x}$. We may say that $\bar{F}$ counts the number of successes in the run of trials $\bar{x}$. Let us compute the distribution $\bar{Q}$ of $\bar{F}$.

37° Let j be any member of $\bar{Y}$. Of course, j is an integer for which $0 \leq j \leq n$. Let $\bar{x}$ be any member of $\bar{X}$. Obviously, $F(\bar{x}) = j$ iff the number of 1s in $\bar{x}$ equals j. By our previous discussion of the Selection Principle and the Placement Principle, we know that there are:

$$\binom{n}{j}$$

such $\bar{x}$. Moreover, for any such $\bar{x}$, we have:

$$p_{\bar{x}} = p_0^{n-j} \times p_1^j$$

We conclude that:

$$\bar{Q}(j) = \binom{n}{j} \times p_0^{n-j} \times p_1^j$$

38• Let $p_0 = p_1 = 1/2$ and let $n = 12$. Compute the binomial distribution on the set:

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$$

Compare your result with the result of article $32^\bullet$.

39° Imagine a coin. Let the coin be weighted in such a way that, when flipped, the coin shows tails (0) with probability $p_0 = 1/3$ and heads (1) with probability $p_1 = 2/3$. Imagine a croupier (let him be Bernoulli). Let the croupier flip the coin 15 times in succession, "independently." Calculate the probability that heads shows on 10 of the flips.

40° Very often, a random variable takes its values in the set $\mathbf{R}$ of real numbers. That was true for the random variable H defined on the magnetic string space with twelve atoms (see article $32^\bullet$), for the random variable F defined on the dice roll space (see article $33^\bullet$), and for the various random variables $\bar{F}$ defined on the various bernoulli n-trial spaces just described. For such a random variable, we can define a particular real number: the *expected value*, which provides useful information.

Let X be a finite probability space with probability function P defined on the subsets of X. Let Y be a (finite) subset of $\mathbf{R}$ and let F be a random variable defined on X with values in Y. One defines the *expected value* of F as follows:

$$E(F) := \sum_{x \in X} F(x) \times P(x)$$

Of course, for cases in which the elementary events in X are equally likely, we have:

$$E(F) = \frac{1}{n} \times \sum_{x \in X} F(x)$$

which is the *average* value of F. We may say that idea of expected value is a generalization of the idea of average value.

$41^\bullet$ Show that:

$$E(F) = \sum_{y \in Y} y \times Q(y)$$

where Q is the distribution of F.

$42^\bullet$ In context of the magnetic string space with twelve atoms, show (deftly) that:

$$E(H) = 0$$

$43^\bullet$ Let n be a positive integer. Let p_0 and p_1 be real numbers for which $0 \leq p_0 \leq 1$, $0 \leq p_1 \leq 1$, and $p_0 + p_1 = 1$. Let $\bar{X}$ be the corresponding bernoulli n-trial space with probability function $\bar{P}$ defined on the subsets of $\bar{X}$. Let $\bar{F}$ be the random variable defined on $\bar{X}$ which "counts the number of successes." Show that:

$$E(\bar{F}) = n \times p_1$$

Let us write out the solution to this problem. To do so, we will break our promise in article 1° by applying a trace of the Calculus.

Consider the binomial expansion:

$$(p_0 + p_1)^n = \sum_{j=0}^{n} \binom{n}{j} \times p_0^{n-j} \times p_1^j$$

where p_0 and p_1 are for now any real numbers. One can justify the foregoing expansion by referring to our discussion of counting principles in section 3. Imagine p_0 to be constant. Take the derivative with respect to p_1:

$$n \times (p_0 + p_1)^{n-1} = \sum_{j=0}^{n} j \times \binom{n}{j} \times p_0^{n-j} \times p_1^{j-1}$$

Now consider p_0 and p_1 to be constrained as follows:

$$0 \leq p_1 \leq 1, \quad p_0 = 1 - p_1$$

Multiply by p_1. Explain why we are finished.

5 Conditioned Probability

44° Let X be a finite probability space and let P be the associated probability function defined on the subsets of X. Let Y and Z be any finite sets and let F and G be random variables defined on X with values in Y and Z. Let Q and R be the corresponding distributions of F and G. Let us imagine that Tyche draws an elementary event, let it be x, at random from X. Let $y := F(x)$ and $z := G(x)$. Of course, we may proceed to calculate $Q(y)$ and $R(z)$.

In this section, we will consider a rather subtle and important question. Given foreknowledge of the value z, how shall we modify our computation of $Q(y)$?

45° For an example, let us return to the dice roll space X with equally likely outcomes. The elementary events are the following:

$$\begin{array}{cccccc}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\
(6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)
\end{array}$$

and the corresponding probabilities all equal 1/36. Let Y be the finite set composed of the integers:

$$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$$

and let Z be the finite set composed of the integers:

$$1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 18, 20, 24, 25, 30, 36$$

Let F and G be the random variables defined on X with values in Y and Z, as follows:

$$F(j,k) := j + k$$
$$G(j,k) := j \times k$$

where (j,k) is any elementary event in X. For a given roll of the dice, we might ask for the probability that the sum of the faces which come up equals 7, given foreknowledge that the product of the faces which come up equals 12. Common sense suggests this "conditioned" probability should be (not the previously determined, unconditioned probability 1/6) but 1/2. Why?

46° To make the idea of conditioned probability precise requires effort. Let us proceed carefully.

Let y be any member of Y and let z be any member of Z. With reference to article 30°, let $F^{-1}(y)$ be the subset of X composed of all members x for which $F(x) = y$ and let $G^{-1}(z)$ be the subset of X composed of all members x for which $G(x) = z$. By definition, we have:

$$Q(y) = P(F^{-1}(y))$$

and:

$$R(z) = P(G^{-1}(z))$$

47° Let us assume now that the value of G is "foreknown" to be z. It follows that x, however Tyche may choose it at random, must lie in $G^{-1}(z)$. We are led to define a new probability function P_z on the subsets of X, as follows:

$$P_z(A) := \frac{P(A \cap G^{-1}(z))}{R(z)}$$

where A is any subset of X. We obtain a new probability function Q_z defined on the subsets of Y, namely, the distribution of F relative not to P but to P_z:

$$Q_z(B) := P_z(F^{-1}(B))$$

where B is any subset of Y. In particular, for each y in Y:

$$Q_z(y) := P_z(F^{-1}(y))$$

We refer to $Q_z(y)$ as the *conditioned* probability that the value of F is y, given that the value of G is z.

48° Similarly, let us assume that the value of F is "foreknown" to be y. It follows that x, however Tyche may choose it at random, must lie in $F^{-1}(y)$. We are led to define a new probability function P_y on the subsets of X, as follows:

$$P_y(A) := \frac{P(F^{-1}(y) \cap A)}{Q(y)}$$

where A is any subset of X. We obtain a new probability function R_y defined on the subsets of Z, namely, the distribution of G relative not to P but to P_y:

$$R_y(C) := P_y(G^{-1}(C))$$

where C is any subset of Z. In particular, for each z in Z:

$$R_y(z) := P_y(G^{-1}(z))$$

We refer to $R_y(z)$ as the *conditioned* probability that the value of G is z, given that the value of F is y.

49° Let us give the foregoing definitions a somewhat more pleasing form. For any y in Y and for any z in Z, let:

$$\Pi(y, z) := P(F^{-1}(y) \cap G^{-1}(z))$$

Clearly, $\Pi(y, z)$ is the probability that, for an elementary event x in X chosen by Tyche at random, the values $F(x)$ and $G(x)$ equal y and z.

50$^\bullet$ Verify that, for any y in Y:

$$Q(y) = \sum_{z \in Z} \Pi(y, z)$$

and that, for any z in Z:

$$R(z) = \sum_{y \in Y} \Pi(y, z)$$

51° By inspection, we find that:

$$Q_z(y) = \frac{\Pi(y,z)}{R(z)}$$

and that:

$$R_y(z) = \frac{\Pi(y,z)}{Q(y)}$$

52$^\bullet$ Apply the foregoing discussion to make precise the common sense conclusion in article 45$^\bullet$.

53° In context of the foregoing discussion, let us note that, for any y in Y and for any z in Z:

$$Q_z(y) \times R(z) = \Pi(y,z) = R_y(z) \times Q(y)$$

By this relation and by article 50$^\bullet$, we infer that, for each z in Z, the array:

$$R_y(z), \;\; Q(y) \qquad (y \in Y)$$

determines the array:

$$\Pi(y,z) \qquad (y \in Y)$$

hence the array:

$$R(z), \;\; Q_z(y) \qquad (y \in Y)$$

Similarly, for each y in Y, the array:

$$Q_z(y), \;\; R(z) \qquad (z \in Z)$$

determines the array:

$$\Pi(y,z) \qquad (z \in Z)$$

hence the array:

$$Q(y), \;\; R_y(z) \qquad (z \in Z)$$

These observations comprise the celebrated Theorem of Bayes.

54° It may happen that foreknowledge of G has no effect upon the distribution of F and that foreknowledge of F has no effect upon the distribution of G. In such a case, we say that F and G are *independent*. Clearly, F and G are independent iff, for each y in Y and for each z in Z:

$$Q_z(y) = Q(y) \quad \text{and} \quad R_y(z) = R(z)$$

which is to say that, for each y in Y and for each z in Z:

$$\Pi(y, z) = Q(y)R(z)$$

55$^\bullet$ Determine whether or not the random variables F and G in article 45° are independent.

56$^\bullet$ Would you expect that, for an individual chosen at random, height and weight would be independent? Height and IQ?

57$^\bullet$ Let D be a dread disease and let T be a test by which one presumes to determine whether or not a given individual has the disease.

Let Y and Z both equal the set composed of the members:

$$0,\ 1$$

and let $X := Y \times Z$. We may display the members of X in the usual fashion:

$$\begin{matrix} (0,1) & (1,1) \\ (0,0) & (1,0) \end{matrix}$$

Let F and G be the mappings carrying X to Y and Z, defined as follows:

$$F(j,k) := j$$
$$G(j,k) := k$$

where (j,k) is any member of X.

For an individual chosen at random, we imagine that precisely one of the four "labels" in X applies. Thus, $(0,0)$ applies if the individual does not have the disease and tests negative; the label $(1,0)$ applies if the individual does have the disease but tests negative; and so forth. Let the probabilities for these "elementary events" be the following:

$$\begin{matrix} P(0,1) & P(1,1) \\ P(0,0) & P(1,0) \end{matrix}$$

These probabilities determine the probability function P defined on the subsets of X. Let Q and R be the corresponding distributions of F and G.

Now let us assume that $Q(0) = 0.98$, $Q(1) = 0.02$, $R_0(1) = 0.04$, and $R_1(1) = 0.90$. Apply the Theorem of Bayes to compute $Q_1(0)$. What do these probabilities "mean"?

Let us write out the solution of this problem.

From $Q(0) = 0.98$, we infer that, for an individual chosen at random, the probability that he does not have the disease is 0.98. From $Q(1) = 0.02$, we infer that, for an individual chosen at random, the probability that he does have the disease is 0.02. Of course, $Q(0) + Q(1) = 1$. From $R_0(1) = 0.04$, we infer that, for an individual chosen at random, the probability that he tests positive, given that he does not have the disease, is 0.04. From $R_1(1) = 0.90$, we infer that, for an individual chosen at random, the probability that he tests positive, given that he does have the disease, is 0.90.

For an individual chosen at random, the probability that he does not have the disease, given that he tests positive, is $Q_1(0)$. That is, $Q_1(0)$ is the probability of a *false positive*, every patient's nightmare.

Let us compute $Q_1(0)$. By articles $50^\bullet$ and 53°, we have:

$$\begin{aligned}\Pi(0,1) &= R_0(1) \times Q(0) = 0.04 \times 0.98 = 0.0392\\ \Pi(1,1) &= R_1(1) \times Q(1) = 0.90 \times 0.02 = 0.0180\end{aligned}$$

hence:

$$R(1) = \Pi(0,1) + \Pi(1,1) = 0.0392 + 0.0180 = 0.0572$$

and therefore:

$$Q_1(0) = \frac{\Pi(0,1)}{R(1)} = \frac{0.0392}{0.0572} \approx 0.6853$$

The given data are very plausible but the probability of a false positive is very high. One should bear this point in mind.

Just as well, we can compute the probability $Q_0(1)$ of a *false negative*:

$$\begin{aligned}\Pi(0,0) &= R_0(0) \times Q(0) = 0.96 \times 0.98 = 0.9408\\ \Pi(1,0) &= R_1(0) \times Q(1) = 0.10 \times 0.02 = 0.0020\end{aligned}$$

hence:

$$R(0) = \Pi(0,0) + \Pi(1,0) = 0.0392 + 0.0180 = 0.9428$$

and therefore:

$$Q_0(1) = \frac{\Pi(1,0)}{R(0)} = \frac{0.0020}{0.9428} \approx 0.0021$$

Of course, the consequences of a false negative are potentially far more grave than the consequences of a false positive. In the design of tests, it is important to make the former as small as possible.

6 A Paradox

58° Let us again consider the dice roll space X with equally likely outcomes. The elementary events are the following:

$$\begin{array}{cccccc}(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6)\\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6)\\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6)\\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6)\\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6)\\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)\end{array}$$

and the corresponding probabilities all equal 1/36. For added interest, let us introduce a Croupier, a Gambler, an Observer, and a circular table neatly covered with green felt. The Gambler is blind. Imagine the following conversation among the three characters.

Croupier (to the Gambler):
"The dice are fair. I have just rolled them out upon the table. What is the probability that I rolled a seven?

Gambler:
"There are 6 favorable elementary events and 36 possible elementary events, so the probability that you rolled a seven is 6/36. I might say 2/12, or even 1/6.

Croupier (to the Gambler):
"Again, I have just rolled the dice out upon the table. What is the probability that I rolled a seven?

Observer (quietly intervening, in a whisper):
"I see a 5."

Gambler:
"I cannot deny that I heard the comment. Now there are 2 favorable elementary events and 11 possible elementary events, so the probability that you rolled a seven is 2/11."

Croupier (to the Gambler):
"Again, I have just rolled the dice out upon the table. What is the probability that I rolled a seven?

Observer (quietly intervening, in a whisper):
"I see a 2."

Gambler:
"I heard that. Again there are 2 favorable elementary events and 11 possible elementary events, so the probability that you rolled a seven is 2/11."

Croupier (to the Gambler):
"Again, I have just rolled the dice out upon the table. What is the probability that I rolled a seven?

Observer (quietly intervening, in a whisper):
"I see a ... *cough, cough*."

Gambler:
"I did not hear the comment. But it does not matter. Whatever the comment, there are now 2 favorable elementary events and 11 possible elementary events, so the probability that you rolled a seven is 2/11."

Croupier (to the Gambler):

"Again, I have just rolled the dice out upon the table. What is the probability that I rolled a seven?

Observer (quietly intervening, in a whisper):
"No comment."

Gambler (hesitating):
"There seems to be trouble here."

7 A Case Study: Woven Fabrics

59° In this final section, we will describe an interesting problem which arises in the study of woven fabrics. The problem can be expressed in the terms of probability theory. One may try to solve the problem by computing or by thinking. To this date, the problem is unsolved: the required computation is massive and the required ratiocination is elusive.

One should consult the brief readable account of the problem in the article:

> "When a fabric hangs together"
> Bull. London Math. Soc., 12 (1980) 161-164

by C. R. J. Clapham.

60° We may represent a woven fabric as a rectangular array of zeros and ones, such as the following:

$$F: \quad \begin{matrix} 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{matrix}$$

The rows in the array represent the weft (horizontal) threads in the fabric and the columns represent the warp (vertical) threads. For each weft thread and for each warp thread, the number in the array standing at the intersection of the two threads indicates which of the two lies over the other. In particular, 0 indicates that the weft thread lies over the warp thread and 1 indicates that the warp thread lies over the weft thread.

61° Now let m and n be any positive integers. Let us imagine a fabric F having m rows (that is, weft threads) and n columns (that is, warp threads).

We say that F *falls apart* iff it is possible to mark certain of the rows in F and certain of the columns in F so that each of the marked rows lies over all of the unmarked columns and so that each of the marked columns lies over all of the unmarked rows. In such a case, the marked rows and columns will "lift away" from the unmarked rows and columns, with the result that the fabric falls apart. Of course, we understand that the number p of the marked rows and the number q of the marked columns meet the conditions $0 \leq p \leq m$, $0 \leq q \leq n$, and $0 < p + q < m + n$.

62$^\bullet$ Show that the fabric displayed in article 60° falls apart. To that end, start with the second column and show that the first, third, fifth, and seventh rows and the second, fourth, sixth, and eighth columns lift away from the rest.

63° Let:

$$M := \{1, 2, 3, \ \ldots \ , m\}$$

and let:

$$N := \{1, 2, 3, \ \ldots \ , n\}$$

For each k in M and for each ℓ in N, let:

$$F_{k\ell}$$

stand for the entry in F standing in the k-th row and the ℓ-th column. For each k in M, let:

$$R_k := \sum_{\ell \in N} F_{k\ell}$$

be the corresponding row sum, and, for each ℓ in N, let:

$$C_\ell := \sum_{k \in M} F_{k\ell}$$

be the corresponding column sum.

Now let P be any subset of M and let Q be any subset of N. Let p be the number of members of P and let q be the number of members of Q. Let the rows corresponding to the members of P be marked and let the columns corresponding to the members of Q be marked. We contend that the marked rows and columns lift away from the rest iff:

$$\sum_{\ell \in Q} C_\ell \ - \ \sum_{k \in P} R_k \ = \ q(m - p)$$

64$^\bullet$ Prove the foregoing contention. To do so, introduce:

$$G := \sum_{k \in P} \sum_{\ell \in Q} F_{k\ell}$$

Then show that:

$$\sum_{\ell \in Q} C_\ell \;\leq\; q(m-p) + G \;\leq\; q(m-p) + \sum_{k \in P} R_k$$

Finally, note that the marked rows and columns lift away from the rest iff:

$$G \;=\; \sum_{k \in P} R_k$$

and:

$$\sum_{\ell \in Q} C_\ell \;=\; q(m-p) + G$$

65° We say that the fabric F *hangs together* iff it does not fall apart. For the weaver, that is a desirable property of a fabric.

66• Let m and n be any positive integers. Let $\mathcal{F}$ be the family of all possible fabrics F having m rows and n columns. Note that there are:

$$2^{m \times n}$$

members of $\mathcal{F}$. Imagine a weaver who weaves at random, producing the various possible fabrics F with equal probability. What is the probability that the weaver produces a fabric which hangs together?