The Sample Space.

Definition. *Sample Point.* Every thinkable outcome of a random experiment is completely described by one, and only one sample point.

Definition. *Event.* We define the word event to mean the same as an aggregate of sample points. We shall say that an event A consists of (or contains) certain points, namely those representing outcomes of the ideal experiment in which A occurs.

Distribution of r balls into n cells.

The below table describes all possible outcomes of the experiment of placing three balls into three cells.

1.	abc	-	-
2.	-	abc	-
3.	-	-	abc
4.	ab	c	-
5.	ac	b	-
6.	bc	a	-
7.	ab	-	c
8.	ac	-	b
9.	bc	-	a

10.	a	bc	-
11.	b	ac	-
12.	c	ab	-
13.	a	-	bc
14.	b	-	ac
15.	c	-	ab
16.	-	ab	c
17.	-	ac	b
18.	-	bc	a

19.	-	a	bc
20.	-	b	ca
21.	-	c	ab
22.	a	b	c
23.	a	c	b
24.	b	a	c
25.	b	c	a
26.	c	a	b
27.	c	b	a

The sample space consists of $3^3 = 27$ points. Each of the arrangements represents a simple event, that is, a sample point.

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 description.

(b,1) *Birthdays.* The possible configurations of the birthdays of r people correspond to the different arrangements of r balls into $n = 365$ cells.

(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) In *firing* at n targets, the hits correspond to the balls, the targets to the cells.

(b,4) *Sampling.* Let a group of r people be classified according to, say, age or profession.

The classes play the role of our cells, the people that 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 the study of the genetic effect of irradiation, the chromosomes correspond to the cells of our model and α-particles to the balls.

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

(b,7) *Dice.* The possible outcomes of a throw with r dice correspond to placing r balls into $n = 6$ cells. When *tossing a coin* we are in effect dealing with $n = 2$ cells.

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

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

(b,11) *Coupon collecting.* The different kind of coupons represent the cells, the coupons collected represent the balls.

(b,12) *Aces in a bridge.* The four players represent four cells, and we have $r = 4$ balls.

(b,13) *Gene distribution.* Each descendant of an individual, person, plant or animal inherits from the progenitor certain genes. If a particular gene can appear in n forms $A_1, A_2, \ldots, A_n$, then the descendants may be classified according to the type of the gene. The descendants correspond to the balls, the genotypes $A_1, A_2, \ldots, A_n$ to the cells.

(b,14) *Chemistry.* Suppose that 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 the balls and the polymer chains the role of the cells into which the 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 r, of quanta. For the theory of black-white contrast we must know, how many cells are likely to be hit by the r 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 more complicated since a plate usually contains grains of different sensitivity).

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

The case of indistinguishable balls.

Consider the experiment of distribution $r = 3$ balls into $n = 3$ cells. Suppose that the balls are no longer distinguishable. This means that we no longer distinguish between three arrangements such as $4, 5, 6$. The sample space of the experiment which we call, *placing three indistinguishable balls into three cells* is as follows:

1.	***	-	-
2.	-	***	-
3.	-	-	***
4.	**	*	-
5.	**	-	*

1.	*	**	-
2.	*	-	**
3.	-	**	*
4.	-	*	**
5.	*	*	*

In the scheme above, we have considered indistinguishable balls, but table 2 still refers to a first, second and third cell and their order is indistinguishable.

Relation among events.

Definition. *Union and intersection of events.* To every collection $A, B, C, \ldots$ of events, we define two new events as follows. The aggregate of the sample points which belong to all the given sets will be denotes by $ABC \ldots$ and called the intersections (or simultaneous realization) of $A, B, C, \ldots$. The aggregate of sample points which belong to atleast on of the given sets will be denoted by $A \cup B \cup C \ldots$ and called the union (or realization of atleast one) of the given events. The events $A, B, C, \ldots$ are mutually exclusive if no two have a point in common, that is, if $AB = 0, AC = 0, \ldots, BC = 0, \ldots$.

Definition. *Event B implies A.* The symbols $A \subset B$ and $B \subset A$ are equivalent and signify that every point of A is contained in B, they are read, respectively A implies B and B is implied by A. If this is the case, we shall also write $B - A$ instead of BA^C to denote the event that B but not A occurs.

Discrete Sample Spaces.

Definition. *Discrete Sample Space.* A sample space is called discrete if it contains only finitely many points or infinitely many points which can be arranged into a simple sequence E_1, E_2, ...

We are confronted with a distinction familiar in mechanics. There it is usual first to consider discrete mass points with each individual point carrying a finite mass, and then to pass to the notion of a continuous mass distribution, where each individual point has zero mass. In the first case, the mass of a system is obtained simply by adding the masses of the individual points; in the second case, masses are computed by integration over mass densities. Quite similarly, the probabilities of events in discrete sample spaces are obtained by mere additions, whereas in other spaces integrations are necessary. Except for the technical tools required, there is no essential difference between the two cases. In order to present actual probability considerations unhampered by technical difficulties, we shall take up only discrete sample spaces. It will be seen that even this special case leads to many interesting and important result.

Probabilities in Discrete Sample Spaces: Preparations

The probabilities of the various events are numbers of the same nature as distances in geometry or masses in mechanics. The theory assumes hat they are given, but need assume nothing about their actual numerical values or how they are measured in practice. Some of the most important applications are of a qualitative nature and independent of numerical values; the general conclusions of the theory are applied in many ways exactly as the theorems of geometry serve as a basis for physical theories and engineering applications.

Examples

(a) *Distinguishable balls.* In the experiment of placing $r = 3$ balls into $n = 3$ cells, it appears natural to assume that all sample points are equally probable, that is each sample point has probability $\frac{1}{27}$. We can start from this definition and investigate its consequences. Whether or not out model will come reasonably close to actual experience will depend on the type of phenomena to which it is applied. In some applications the assumption of equal probabilities is imposed by physical considerations; in others it is introduced to serve as the simplest model for a general orientation, even though it quite obviously represents only a crude first approximation.

(b) *Indistinguishable balls: Bose-Einstein statistics.* We now turn to the example of placing three indistinguishable balls in three cells. It is possible to argue that the actual physical experiment is unaffected by our failure to distinguish between the balls; physically there remain 27 different possibilities even though only ten different forms are distinguishable. This consideration leads us to attribute the following probabilities to the ten points of the below table.

Point Number:	1	2	3	4	5	6	7	8	9	10
Probability:	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{2}{9}$

Historically, our argument was accepted for a long time without question and served as the basis for the derivation of the *Maxwell-Boltzmann statistics* for the distribution of r balls in n cells. The greater was the general surprise when Bose and Einstein showed that certain particles are subject to *Bose-Einstein* statistics. In our case, with each $r = n = 3$, the Bose-Einstein model attributes $\frac{1}{10}$ to each of the ten sample points.

The Basic Definitions and Rules.

Fundamental Convention. Given a discrete sample space Ω with sample points $\omega_1, \omega_2, \ldots$, we shall assume that with each point ω_j there is associated a number, called the probability of ω_j and denoted by $P(\omega_j)$. It is to be non-negative and such that

$$P(\omega_1) + P(\omega_2) + \ldots = 1 \tag{1}$$

Definition. *Probability of an Event.* The probability $P(A)$ of any event A is the sum of the probabilities of all sample points in it.

The fundamental equation (1) states that the probability of the entire sample space Ω is unity, or $P(\Omega) = 1$. It follows that for any event A

$$0 \leq P(A) \leq 1 \tag{2}$$

Consider now two arbitrary events A_1 and A_2. To copute the probability $P(A_1 \cup A_2)$ that either A_1 or A_2 or both occur, we have to add the probabilities of all sample point contained either in A_1 or A_2, but each point is to be counted only once. We have, therefore,

$$P(A_1 \cup A_2) \leq P(A_1) + P(A_2) \tag{3}$$

Now, if E is any point contained both in A_1 and in A_2, then $P(E)$ occurs twice in the right-hand member but only once in the left-hand member. Therefore, the right side exceeds the left side by the amount $P(A_1 A_2)$, and we have the simple but important

Theorem. *Probability of atleast one event.* For any two events A_1 and A_2, the probability that either A_1 or A_2 or both occur is given by

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 A_2) \tag{4}$$

If $A_1 A_2 = 0$, that is if A_1 and A_2 are mutually exclusive, then the equation reduces to

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) \tag{5}$$

The probability $P(A_1 \cup A_2 \cup A_3 \cup \ldots)$ of the realization of at least one among n events can be computed by a formula analogous to the equation above; this will be taken up in chapter IV, section I. Here we note only that the inequality $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$ holds in general. Thus, for arbitrary events, $A_1, A_2, \ldots$, the inequality

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots \tag{6}$$

Occasionally inequality (6) is referred to as Boole's inequality.

Exercise Problems.

1. [Feller, 1.8.1] Among the digits $1, 2, 3, 4, 5$, the first one is chosen and then a second selection is made among the remaining four digits. Assume that all twenty possible results have the same probability. Find the probability that an odd digit will be selected (a) the first time (b) the second time (c) both times.

Solution.

Let A_1 be the event that the first digit is odd, A_2 be the event that the second digit is odd.

$$\begin{aligned}
P(A_1) &= \frac{3 \cdot 4}{20} \\
P(A_2) &= \frac{3 \cdot 4}{20} \\
P(A_1 A_2) &= \frac{3 \cdot 2}{20} \\
P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1 A_2) \\
&= \frac{12}{20} + \frac{12}{20} - \frac{6}{20} \\
&= \frac{18}{20} = \frac{9}{10}
\end{aligned}$$

2. [Feller 1.8.3] Consider the 24 possible arrangements (permutations) of the symbols 1234 and attach to each probability $\frac{1}{24}$. Let A_i be the event that the digit i appears at its natural place (where $i = 1, 2, 3, 4$). Verify formula (4).

Solution. Enumerating the sample space, we have:

$$\begin{aligned}\Omega = \{&1234, 1243, 1324, 1342, 1423, 1432,\\ &2134, 2143, 2314, 2341, 2413, 2431,\\ &3124, 3142, 3214, 3241, 3412, 3421,\\ &4123, 4132, 4213, 4231, 4312, 4321\}\end{aligned}$$

$$\begin{aligned}P(A_1) &= \frac{3\cdot 2\cdot 1}{24} = \frac{6}{24}\\ P(A_2) &= \frac{6}{24}\\ P(A_1 \cup A_2) &= \frac{10}{24}\\ P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1A_2)\\ &= \frac{6}{24} + \frac{6}{24} - \frac{2}{24} = \frac{10}{24}\end{aligned}$$

3. [Feller, 1.8.4] A coin is tossed until for the first time the same result appears twice in succession. To every possible outcome requiring n tosses attribute probability $1/2^n$. Describe the sample space. Find the probability of the following events: (a) the experiment ends before the sixth toss, (b) an even number of tosses is required.

Solution.

(a) Enumerating the sample space, we have:

HH	TT
HTT	THH
HTHH	THTT
HTHTT	THTHH
HTHTHH	THTHTT
HTHTHTT	THTHTHH
$\vdots$	$\vdots$

The probability that the experiment ends before $n = 6$ tosses is $2\left(\frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \frac{1}{2^5}\right) = \frac{15}{16}$.

The probability that an even number of tosses are required is $\frac{2}{2^2} + \frac{2}{2^4} + \ldots$. The sum of this infinite series is:

$$\begin{aligned} S &= \frac{1}{2}\left(1 + \frac{1}{2^2} + \frac{1}{2^4} + \ldots\right) \\ &= \frac{1}{2} \cdot \frac{1}{1 - \frac{1}{2^2}} \\ &= \frac{1}{2} \cdot \frac{4}{3} \\ &= \frac{2}{3} \end{aligned}$$

4. [Feller, 1.8.9] Two dice are thrown. Let A be the event that the sum of the faces is odd, B be the event of at least one ace. Describe the events $AB, A \cup B, AC^C$. Find their probabilities assuming that all 36 sample points have equal probabilities.

Solution.