

Study Notes and Exercise Solutions

An Introduction of Probability Theory and Its Applications

Jianyu Zhou

December 12, 2014

Contents

1	The Sample Space	2
1.1	The Empirical Background	2
1.2	Examples	2
1.3	The Sample Space and Events	2
1.4	Relations among Events	3
1.5	Discrete Sample Spaces	4
1.6	Probabilities in Discrete Sample Spaces: Preparations	4
1.7	The Basic Definitions and Rules	4
1.8	Problems Solution	6
2	Elements of Combinatorial Analysis	10
2.1	Preliminary	10
2.2	Ordered Samples	10
2.3	Examples	11
2.4	Subpopulations and partitions	12
2.5	Application to Occupancy Problems	14
3	Fluctuations in Coin Tossing and Random Walks	15
4	Combination of Events	16
5	Conditional Probability, Stochastic Independence	17

1 The Sample Space

1.1 The Empirical Background

Events Events are the results of experiments or observations. The events should be distinguished between *compound* (*decomposable*) and *simple* (*indecomposable events*). A compound event is an aggregate of certain simple events.

Sample points Sample points are just the simple events. Every indecomposable result of the (idealized) experiment is represented by one, and only one, sample point.

Sample Space Sample space is the aggregate of all sample points.

1.2 Examples

Notes

1. All intuitive background sampling problems are abstractly equivalent to the scheme of **placing r balls into n cells**, in the sense that the outcomes differ only in their verbal description.
(There are some examples listed on P10.)
2. Here comes a question that how to calculate the size of sampling space (how many sample points in the sample space) when placing r balls into n cells? This question should be answered under different situations.
 - Placing r **distinguishable** balls into n **distinguishable** cells.

$$N = n^r$$

- Placing r **indistinguishable** balls into n **distinguishable** cells.

$$N = \binom{r+n-1}{r}$$

When we facing some particular problems, the model of distinguishable or indistinguishable balls is purely a matter of purpose and convenience.

1.3 The Sample Space and Events

Sample space is the universe of the sample points. All sample points are indecomposable and cover all the outcomes of an idealized experiment. The event is an aggregate of some sample points (one or the aggregate of all sample points is also an event).

New events can be define in terms of two or more given events. Here comes the notation of the formal *algebra of events* (algebra of point sets).

1.4 Relations among Events

We use Ω to denote sample space and capitals to denote events (sets of sample points). The fact that a sample point x is contained in the event A is denoted by $x \in A$. Thus $x \in \Omega$ for every sample point x . We write $A = B$ only if the two events consist of exactly the same points.

Definitions

1. We shall use the notation $A = 0$ to express that the event A contains no sample points. The zero must be interpreted in **a symbolic sense and not as the numeral**.
2. The event consisting of all points not contained in the event A will be called the *complementary event* (or *negation*) and will be denoted by A' . In particular $\Omega' = 0$.
3. With any two events A and B we can associate two new events defined by the conditions “*both A and B occur*” and “*either A or B or both occur*”. These events will be denoted by AB (intersection of A and B) and $A \cup B$ (union of A and B). If A and B exclude each other, then there are no points common to A and B and the event AB is impossible; analytically this situation is described by the equation $AB = 0$ which should be read “*A and B are mutually exclusive*”.
4. To every collection A, B, C, \dots of events we define some notions as follows.
 - The aggregate of sample points which belong to all the given sets will be denoted by $ABC \dots$ and called the *intersection* (or *simultaneous realization*).
 - The aggregate of sample points which belong to at least one of the given sets will be denoted by $A \cup B \cup C \cup \dots$ and called the *union* (or *realization of at least one*).
 - The events A, B, C, \dots are *mutually exclusive* if no two have a point in common (any two of the events are mutually exclusive), that is, if $AB = 0, AC = 0, \dots, BC = 0, \dots$.
5. The symbols $A \subset B$ and $B \supset A$ 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' to denote the event that *B but not A occurs*. The event $B - A$ contains all those points which are in B but not in A . With this notation we can write $A' = \Omega - A$ and $A - A = 0$.
 - If A and B are mutually exclusive, then the occurrence of A implies the non-occurrence of B and vice versa. Thus $AB = 0$ means the same as $A \subset B'$ and as $B \subset A'$.
 - The event $A - AB$ means the occurrence of A but not of the A and B . Thus $A - AB = AB'$.

Result	Probability	Result	Probability
$\{** * - -\}$	$\frac{1}{27}$	$\{* * * -\}$	$\frac{1}{9}$
$\{- * * * -\}$	$\frac{1}{27}$	$\{* - * *\}$	$\frac{1}{9}$
$\{- - ***\}$	$\frac{1}{27}$	$\{- * * *\}$	$\frac{1}{9}$
$\{* * * -\}$	$\frac{1}{9}$	$\{- * * *\}$	$\frac{1}{9}$
$\{* * - *\}$	$\frac{1}{9}$	$\{* * *\}$	$\frac{2}{9}$

Table 1: The probability distribution

1.5 Discrete Sample Spaces

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, \dots .

The probabilities of events in discrete sample spaces are obtained by mere additions, whereas in other spaces integrations are necessary.

1.6 Probabilities in Discrete Sample Spaces: Preparations

The notion of probability is based on an *idealized model*. In a sense, no ideal models exist in reality. But in many applications, it is sufficiently accurate to describe reality and the idealized model can be extremely useful when the experiment is impossible to reconstruct.

Here comes some interesting examples. When placing 3 distinguishable balls into 3 cells, there will be 27 distinguishable possibilities. It appears natural to assume that all sample points are *equally probable*, that is, that each sample point has probability $\frac{1}{27}$. If the 3 balls are indistinguishable, there remain 27 different possibilities, even though only ten different forms are distinguishable. This consideration leads us to attribute the following probabilities to this ten points as following Table (1). However, in our case with $r = n = 3$, Bose and Einstein showed that certain particles attribute equal probability $\frac{1}{10}$ to each of the ten sample points.

These examples teach us not to **rely too much on a priori arguments and to be prepared to accept new and unforeseen schemes.**

1.7 The Basic Definitions and Rules

Fundamental Convention Given a discrete sample space Ω with sample points E_1, E_2, \dots , we shall assume that with each point E_j there is associated

a number, called the *probability* of E_j and denoted by $P\{E_j\}$. It is to be non-negative and such that

$$P\{E_1\} + P\{E_2\} + \cdots = 1. \quad (1.7.1)$$

Note that the point which probability is zero (which is impossible to occur) is not excluded. In discrete sample spaces, any sample point can be eliminated from the sample space. However, the numerical values of probabilities are not known in advance.

Definition The probability $P\{A\}$ of any event A is the sum of the probabilities of all sample point in it.

For the sample space,

$$P\{\Omega\} = 1. \quad (1.7.2)$$

It follows that for any event A ,

$$0 \leq P\{A\} \leq 1. \quad (1.7.3)$$

Consider now two arbitrary events A_1 and A_2 . To compute 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 points contained either in A_1 or in A_2 , but each point is to be counted only once. Therefore,

$$P\{A_1 \cup A_2\} \leq P\{A_1\} + P\{A_2\}. \quad (1.7.4)$$

Theorem For any two events A_1 and A_2 the probability that either A_1 or A_2 or both occur is given by

$$P\{A_1 \cup A_2\} = P\{A_1\} + P\{A_2\} - P\{A_1 A_2\}. \quad (1.7.5)$$

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

$$P\{A_1 \cup A_2\} = P\{A_1\} + P\{A_2\}. \quad (1.7.6)$$

The probability $P\{A_1 \cup A_2 \cup \cdots \cup A_n\}$ of the realization of at least one among n events can be computed by a formula analogous to the Formula (1.7.5) and will be discussed later. For arbitrary events A_1, A_2, \dots the inequality

$$P\{A_1 \cup A_2 \cup \dots\} \leq P\{A_1\} + P\{A_2\} + \dots \quad (1.7.7)$$

holds (Occasionally (1.7.7) referred to as *Boole's inequality*). In the special case where the events A_1, A_2, \dots are mutually exclusive, we have

$$P\{A_1 \cup A_2 \cup \dots\} = P\{A_1\} + P\{A_2\} + \dots \quad (1.7.8)$$

Finally, the probabilities of all sample points in sample spaces should not be same. Although considering them sharing the same probability will simplify the problem, this assumption will be restricted almost entirely the study of games.

1.8 Problems Solution

These problem are very classic and make me more clear about some notions and enhance what I have learn. Pay more attention to problem 4, 5, 6, 14, 15, 16.

1. (a) $P\{A\} = \frac{3}{5}$,
 (b) $P\{B\} = \frac{3}{5} \times \frac{2}{4} + \frac{2}{5} \times \frac{3}{4} = \frac{3}{5}$
 (c) $P\{C\} = \frac{3 \times 2}{20} = \frac{3}{10}$
2. 6. With the number of points of $S_1, S_2, S_1 \cup S_2$ are 12, 12, 18 respectively, the number of points of $S_1 S_2 = 12 + 12 - 18 = 6$.
3. $P\{A_1\} = P\{A_2\} = P\{A_3\} = P\{A_4\} = \frac{3 \times 2 \times 1}{24} = \frac{1}{4}$,
 $P\{A_1 \cup A_2\} = \frac{3 \times 2 \times 1}{24} + \frac{2 \times 2 \times 1}{24} = \frac{5}{12}$, (the aggregate of the points which the first place is 1 and the points which the second place is 2 but the first place is not 1)
 $P\{A_1 A_2\} = \frac{2 \times 1}{24} = \frac{1}{12}$,
 so, $P\{A_1 \cup A_2\} = P\{A_1\} + P\{A_2\} - P\{A_1 A_2\}$ is valid.
 ...
4. The sample space is all of the outcomes tossing the coin for N times, which contains 2^N sample points. Here N is an infinite number. So, let E_n denote the experiment ends after $n(n \geq 2)$ tossing, and

$$P\{E_n\} = \frac{2 \times 2^{N-n}}{2^n \times 2^{N-n}} = \frac{1}{2^{n-1}}. \quad (1.8.1)$$

Note that $E_2, E_3, \dots, E_n, \dots$ are mutually exclusive.

(a) So

$$\begin{aligned} P\{A\} &= P\{E_2 \cup E_3 \cup E_4 \cup E_5\} \\ &= P\{E_2\} + P\{E_3\} + P\{E_4\} + P\{E_5\} \\ &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{15}{16}. \end{aligned} \quad (1.8.2)$$

(b) Similarly, with k is an positive integer,

$$\begin{aligned} P\{B\} &= P\{E_2 \cup E_4 \cup \dots \cup E_{2k} \cup \dots\} \\ &= P\{E_2\} + P\{E_4\} + \dots + P\{E_{2k}\} + \dots \\ &= \frac{1}{2} + \frac{1}{8} + \dots + \frac{1}{2^{2k-1}} + \dots \\ &= \sum_{k=1}^{\infty} \frac{1}{2^{2k-1}} = \lim_{k \rightarrow \infty} \frac{\frac{1}{2} (1 - \frac{1}{4^k})}{1 - \frac{1}{4}} = \frac{2}{3} \end{aligned} \quad (1.8.3)$$

Notes: This solution is not totally correct. Although the answers are correct, the sample space is not corresponding to the description. The sample space should contain the two points HH and TT with probability $\frac{1}{4}$; the two points HTT and TTH with probability $\frac{1}{8}$; and generally two point with probability $\frac{1}{2^n}$ with $n \geq 2$. So, in conclusion, the sample points are the **outcomes** of the experiment and each point will have it **own's** probability.

5. (a) $P\{Unity\} = 2 \times (\frac{1}{4} + \frac{1}{8} + \dots + \frac{1}{2^k} + \dots) = 1$. (Under the description the unity the points (*) is just the sample space.)
 (b) Let A, B, C denote the events that a, b or c wins respectively. Note that if nobody win the game, the loop shall be either $acbcb \dots$ or $bcabca \dots$. So, a, b winning the game is the either one of the points when there are $3n - 1$ or $3n + 1$ characters ($n = 1, 2, 3 \dots$), i.e.

$$\begin{aligned} P\{A\} = P\{B\} &= \frac{1}{4} + \frac{1}{16} + \dots + \frac{1}{2^{3n-1}} + \frac{1}{2^{3n+1}} + \dots \\ &= \sum_{n=1}^{\infty} \left(\frac{1}{2^{3n-1}} + \frac{1}{2^{3n+1}} \right) \\ &= \lim_{n \rightarrow \infty} \frac{\frac{1}{4} \left(1 - \frac{1}{8^n}\right)}{1 - \frac{1}{8}} + \lim_{n \rightarrow \infty} \frac{\frac{1}{16} \left(1 - \frac{1}{8^n}\right)}{1 - \frac{1}{8}} \\ &= \frac{5}{14} \end{aligned} \tag{1.8.4}$$

and c winning the game is the both points when there are $3n$ characters ($n = 1, 2, 3 \dots$), i.e.

$$\begin{aligned} P\{C\} &= 2 \times \left(\frac{1}{8} + \frac{1}{64} + \dots + \frac{1}{2^{3n}} + \dots \right) \\ &= \sum_{n=1}^{\infty} \frac{2}{2^{3n}} \\ &= \lim_{n \rightarrow \infty} \frac{\frac{2}{8} \left(1 - \frac{1}{8^n}\right)}{1 - \frac{1}{8}} = \frac{2}{7} \end{aligned} \tag{1.8.5}$$

- (c) If no decision is reached at or before the k th turn, the game must be end after the k th turn. So we can consider the complementary event, i.e.

$$\begin{aligned} P\{D\} &= 1 - P\{D'\} \\ &= 1 - 2 \times \left(\frac{1}{2^{k+1}} + \frac{1}{2^{k+2}} + \dots \right) \\ &= 1 - \sum_{n=k}^{\infty} \frac{1}{2^n} \\ &= \lim_{n \rightarrow \infty} \frac{\frac{1}{2^k} \left(1 - \frac{1}{2^n}\right)}{1 - \frac{1}{2}} = \frac{1}{2^{k-1}} \end{aligned} \tag{1.8.6}$$

Note: This is a good problem, and please review this problem in detail. It tells me that before we calculate the probability of the event, we should **determine what the sample space is and the probability distribution of the sample points are in advance**. Then we should **analyze the specific sample points which are included in the event strictly**. Finally, **take more care about the complementary events and whether the events are mutually exclusive**. This is the correct way to calculate the probability of the event.

6. ?
7. It is obvious that $A_1A_2A_3 = A_1A_2A_3A_4 \subset A_4$ and $A_1A_2A'_3 = A_1A_2A'_3A'_4 \subset A'_4$.
8. (a) It is obvious that S_1S_2 and D_3 are mutually exclusive, so $S_1S_2D_3 = 0$.
 (b) Because there are 3 balls, $S_1D_2 = S_1D_2E_3 \subset E_3$ holds.
 (c) There are only 3 balls, so $E_3 = S_1D_2 \cup S_2D_1$ and $(S_1D_2)(S_2D_1) = 0$. Then $E_3 - D_2S_1 \supset S_2D_1$ holds.
9. (a) AB contains the points $(1, 2), (1, 4), (1, 6), (2, 1), (4, 1), (6, 1)$, so $P\{AB\} = \frac{1}{6}$.
 (b) $P\{A \cup B\} = P\{A\} + P\{B\} - P\{AB\} = \frac{1}{2} + \frac{11}{36} - \frac{1}{6} = \frac{23}{36}$.
 (c) $P\{AB'\} = \frac{12}{36} = \frac{1}{3}$.
10. (a) The husband and wife are both older than 40 and the husband is older than wife.
 (b) The husband is older than 40 but younger than wife.
 (c) The husband and wife are both older than 40 and the husband is younger than wife.
11. AC' indicates that the husband is older than 40 but the wife is younger than 40, so the husband must be older than wife and B holds.
12. (a) 0. (b) 0. (c) 4. (d) 2. (e) 1. (f) 1. (g) 0.
13. The verifications are easy to think up considering the description of the problem.
14. The solution can be got by drawing some figures easily. Here I copy these formulas to enhance the impression.
 - (a) $(A \cup B)' = A'B'$
 - (b) $(A \cup B) - B = A - AB = AB'$
 - (c) $AA = A \cup A = A$
 - (d) $(A - AB) \cup B = A \cup B$
 - (e) $(A \cup B) - AB = AB' \cup A'B$
 - (f) $A' \cup B' = (AB)'$
 - (g) $(A \cup B)C = AC \cup BC$

15. (a) $(A \cup B)(A \cup B') = (A \cup B)A \cup (A \cup B)B' = AA \cup BA \cup AB' \cup BB' = A \cup AB' = A$
 (b) $(A \cup B)(A' \cup B)(A \cup B') = B(A \cup B') = BA \cup BB' = AB$
 (c) $(A \cup B)(B \cup C) = AB \cup BB \cup AC \cup BC = B \cup AC \cup BC = B \cup AC$
Notes: $AB \subset A$, $AB \subset B$. Do not confuse.
16. True: (c), (d), (e), (f), (h), (i), (k), (l)
 False:
 (a) The equation will be correct only if $C \subset B$.
 (b) When $AB \neq 0$, $AC = 0$, $BC = 0$ the equation is invalid.
 (g) If $AB \neq 0$, the equation is invalid. It should be $(A \cup B) - A = A'B$.
 (j) Totally wrong. (k) is the correct version.
Notes: In this book, $B - C$ is valid only when $C \subset B$.
17. (a) $AB'C'$
 (b) ABC'
 (c) ABC
 (d) $A \cup B \cup C$
 (e) $AB \cup AC \cup BC$
 (f) $AB'C' \cup A'BC' \cup A'B'C$
 (g) $ABC' \cup AB'C \cup A'BC$
 (h) $A'B'C'$
 (i) $(ABC)'$
18. $A \cup B \cup C = A \cup (B - AB) \cup (C - AC \cup BC) = A \cup A'B \cup A'B'C$
19. A , $B - AB$ and $C - AC \cup BC$ are mutually exclusive, so
- $$\begin{aligned}
 P\{A \cup B \cup C\} &= P\{A\} + P\{A'B\} + P\{A'B'C\} \\
 &= P\{A\} + P\{B\} - P\{AB\} + P\{C\} - P\{AC \cup BC\} \\
 &= P\{A\} + P\{B\} - P\{AB\} + P\{C\} \\
 &\quad - (P\{AC\} + P\{BC\} - P\{ABC\}) \\
 &= P\{A\} + P\{B\} + P\{C\} - P\{AB\} - P\{AC\} - P\{BC\} \\
 &\quad + P\{ABC\}
 \end{aligned}$$
- (1.8.7)

2 Elements of Combinatorial Analysis

The following chapters (chapter 2, 3, 4) are not in conjunction with the chapter 1. These are some complementary knowledge. Chapter 5 will take up the theoretical thread of chapter 1.

2.1 Preliminary

Pairs. With m elements a_1, \dots, a_m and n elements b_1, \dots, b_n , it is possible to form mn pairs (a_j, b_k) containing one element from each group.

Multiplets. Given n_1 elements a_1, \dots, a_{n_1} and n_2 elements b_1, \dots, b_{n_2} , etc., up to n_r elements x_1, \dots, x_{n_r} ; it is possible to form $n_1 n_2 \dots n_r$ ordered r -tuples $(a_{j_1}, b_{j_2}, \dots, x_{j_r})$ containing one element of each kind.

Many applications are based on the following reformulation of the last theorem: r successive selections (decisions) with exactly n_k choices possible at the k th step can produce a total of $n_1 n_2 \dots n_r$ different results.

2.2 Ordered Samples

Ordered Sample. Consider the population (set) of n elements a_1, a_2, \dots, a_n . Any ordered arrangement $a_{j_1}, a_{j_2}, \dots, a_{j_r}$ of r symbols is called an *ordered sample of size r* drawn from the population.

Here two procedures are possible, *sampling with replacement* and *sampling without replacement*.

- **Sampling with replacement.** Here a selection is made from the entire population, so that the same element can be drawn more than once. The samples are then arrangements in which repetitions are permitted. In this case, each of the r element can be chosen in n ways. The number of possible samples is therefore n^r , as can be seen with theorem with $n_1 = n_2 = \dots = n$.
- **Sampling without replacement.** Here an element once chosen is removed from the population, so that the sample becomes an arrangement without repetitions. In this case, r cannot exceed n . We have n possible choices for the first element, but only $n - 1$ for the second, $n - 2$ for the third, etc., and so there are $n(n - 1) \dots (n - r + 1)$ choices in all. This appears so often and it is convenient to introduce the notation

$$(n)_r = n(n - 1) \dots (n - r + 1). \quad (2.2.1)$$

Clearly $(n)_r = 0$ for integers r, n such that $r > n$.

Theorem. For a population of n elements and a prescribed sample size r , there exist n^r different samples with replacement and $(n)_r$ samples without replacement.

We note the special case where $r = n$. n elements a_1, \dots, a_n can be ordered in $(n)_n = n(n-1) \dots 2 \times 1$ different ways. Instead of $(n)_n$ we write $n!$ which is the more usual notion.

Corollary. The number of different ordering of n elements is

$$n! = n(n-1) \dots 2 \times 1. \quad (2.2.2)$$

Whenever we speak of *random samples* of fixed size r , the adjective *random* is to imply that all possible sample points have the *same probability*, namely, n^{-r} in sampling with replacement and $\frac{1}{(n)_r}$ without replacement. If $n \gg r$, the ratio $\frac{(n)_r}{n^r}$ is near unity. This leads us to expect that, for large populations and relatively small samples, two ways of sampling are practically equivalent.

In sampling without replacement the probability for any fixed element of the population to be include in a random sample of size r is

$$1 - \frac{(n-1)_r}{(n)_r} = 1 - \frac{n-r}{n} = \frac{r}{n} \quad (2.2.3)$$

The complementary event is the probability that the fixed element is not chosen. This equation also indicates that no matter when the element will be chosen, the probability for any fixed element being chosen is always $\frac{r}{n}$.

In sampling with replacement the probability that an element be included at least once is

$$1 - \left(1 - \frac{1}{n}\right)^r \quad (2.2.4)$$

2.3 Examples

This section concerns the probability of the event that in the sample no element appears twice from a random sample of size r with replacement is taken from a population of n elements. According to the previous section, there are n^r different samples in all, of which $(n)_r$ satisfy the stipulated condition. Assuming all arrangements have equal probability, the probability is

$$p = \frac{(n)_r}{n^r} \quad (2.3.1)$$

Since sometimes it's hard to calculate the exact value of formula (2.3.1). Here comes two good numerical approximation to p . Here we deduce (2.3.1) into following form.

$$p = \frac{(n)_r}{n^r} = \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{r-1}{n}\right) \quad (2.3.2)$$

When r is small, we can neglect all cross products and have in approximation

$$p \approx 1 - \frac{1 + 2 + \dots + (r-1)}{n} = 1 - \frac{r(r-1)}{2n} \quad (2.3.3)$$

For larger r (still stipulate $r \ll n$) we obtain a much better approximation by passing to logarithms. For small positive x we have $\ln(1-x) \approx -x$, and thus from (2.3.2)

$$\begin{aligned} \ln p &= \ln \left(1 - \frac{1}{n}\right) + \ln \left(1 - \frac{2}{n}\right) + \dots + \ln \left(1 - \frac{r-1}{n}\right) \\ &\approx -\frac{1 + 2 + \dots + (r-1)}{n} = -\frac{r(r-1)}{2n} \end{aligned} \quad (2.3.4)$$

2.4 Subpopulations and partitions

As before, we use the term population of size n to denote an aggregate of n elements without regard to their order. Now consider a subpopulation of size r of a given population consisting of n elements. Since r elements can be numbered in $r!$ different ways, it follows that there are exactly $r!$ times as many samples as there are subpopulations of size r . The number of subpopulations of size r is therefore given by $\frac{(n)_r}{r!}$. Expressions of this kind are known as **binomial coefficients**, and the standard notation for them is

$$\binom{n}{r} = \frac{(n)_r}{r!} = \frac{n(n-1)\dots(n-r+1)}{r(r-1)\dots 2 \cdot 1} \quad (2.4.1)$$

Theorem A population of n elements possesses $\binom{n}{r}$ different subpopulations of size $r \leq n$.

When we choose a subset of r elements, another subset of size $n-r$ is also determined. It follows that there are as many as subpopulations of size r as there are subpopulations of size $n-r$, and hence for $1 \leq r \leq n$ we must have

$$\binom{n}{r} = \binom{n}{n-r} \quad (2.4.2)$$

To prove equation (2.4.2) directly we observe that an alternative way of writing the binomial coefficient (2.4.1) is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \quad (2.4.3)$$

Note that the left side in equation (2.4.2) is not defined for $r = 0$, but the right side is. In order to make equation (2.4.2) valid for all integers r such that $0 \leq r \leq n$, we now define

$$\binom{n}{0} = 1, \quad 0! = 1, \quad (n)_0 = 1. \quad (2.4.4)$$

Stirling's formula can help us calculate the approximation of binomial coefficient, which will be discussed in section 9.

The distinction between distinguishable and indistinguishable elements is similar to the relationship between a subpopulation and the corresponding ordered samples. Deleting the subscripts in an arrangement (or grouping) of r elements a_1, \dots, a_r yields an arrangement of r indistinguishable letters. Different from considering a subpopulation as to **choose r elements from n** , the alternative way to understand this is to **eliminate the repetitive counting of the order information from ordered samples (divide $(n)_r$ by $r!$)**. The latter idea helps us to solve more problems like the examples on P36 (plz review in detail) and lead us to the theorem of partition.

Theorem Let r_1, \dots, r_k be integers such that

$$r_1 + r_2 + \dots + r_k = n, \quad r_i \geq 0 \quad (2.4.5)$$

The number of ways in which a population of n elements can be divided into k ordered parts (partitioned into k subpopulations) of which the first contains r_1 elements, the second r_2 elements, etc., is

$$\frac{n!}{r_1! r_2! \dots r_k!} \quad (2.4.6)$$

This is called **multinomial coefficients**.

Note that the order of the subpopulations is essential in the sense that $(r_1 = 2, r_2 = 3)$ and $(r_1 = 3, r_2 = 2)$ represent different partitions; however, no attention is paid to the order within the groups. It is permitted that $r_i = 0$ because $0! = 1$ makes it no way affect formula (2.4.6).

Proof A repeat use of (2.4.3) will show that (2.4.6) may be rewritten in the form

$$\binom{n}{r_1} \binom{n-r_1}{r_2} \binom{n-r_1-r_2}{r_3} \dots \binom{n-r_1-\dots-r_{k-2}}{r_{k-1}} \quad (2.4.7)$$

On the other hand, in order to effect the desired partition, we have first to select r_1 elements out of the given n ; of the remaining $n - r_1$ elements we select a second group of size r_2 , etc. After forming the $(k-1)$ st group there remain $n - r_1 - r_2 - \dots - r_{k-1} = r_k$ elements and form the last group.

2.5 Application to Occupancy Problems

3 Fluctuations in Coin Tossing and Random Walks

4 Combination of Events

5 Conditional Probability, Stochastic Independence