UNIT-III

Elementary Combinatorics

Basis of counting:

If X is a set, let us use |X| to denote the number of elements in X.

Two Basic Counting Principles

Two elementary principles act as —building blocks< for all counting problems. The first principle says that the whole is the sum of its parts; it is at once immediate and elementary.

Sum Rule: **The principle of disjunctive counting** :

If a set X is the union of disjoint nonempty subsets S1, ….., Sn, then | X | = | S1 | + | S2 | + ….. + | Sn |.

We emphasize that the subsets S1, S2, …., Sn must have no elements in common. Moreover, since X = S1 U S2 U ……U Sn, each element of X is in exactly one of the subsets Si. In other words, S1, S2, …., Sn is a partition of X.

If the subsets S1, S2, …., Sn were allowed to overlap, then a more profound principle will be needed--the principle of inclusion and exclusion.

Frequently, instead of asking for the number of elements in a set *perse*, some problems ask for how many ways a certain event can happen.

The difference is largely in semantics, for if A is an event, we can let X be the set of ways that A can happen and count the number of elements in X. Nevertheless, let us state the sum rule for counting events.

If E1, ……, En are mutually exclusive events, and E1 can happen e1 ways, E2 happen e2 ways,…. ,En can happen en ways, E1 or E2 or …. or En can happen e1 + e2 + …….. + en ways.

Again we emphasize that mutually exclusive events E1 and E2 mean that E1 or E2 can happen but both cannot happen simultaneously.

The sum rule can also be formulated in terms of choices: If an object can be selected from a reservoir in e1 ways and an object can be selected from a separate reservoir in e2 ways and an object can be selected from a separate reservoir in e2 ways, then the selection of one object from either one reservoir or the other can be made in e1 + e2 ways.

Product Rule: The principle of sequencing counting

If S1, ….., Sn ar**e nonempty sets, then the number of elements in the Cartesian product S1 x S2 x ….. x Sn is the product ∏in=1 |S i |. That is,**

| S1 x S2 x .... x Sn | = **∏in=1| S i |.**

Observe that there are 5 branches in the first stage corresponding to the 5 elements of S1 and to each of these branches there are 3 branches in the second stage corresponding to the 3 elements of S2 giving a total of 15 branches altogether. Moreover, the Cartesian product S1 x S2 can be partitioned as (a1 x S2) U (a2 x S2) U (a3 x S2) U (a4 x S2) U (a5 x S2), where (ai x S2) = {(ai, b1), (ai i, b2), (ai, b3)}. Thus, for example, (a3 x S2) corresponds to the third branch in the first stage followed by each of the 3 branches in the second stage.

More generally, if a1,….., an are the n distinct elements of S1 and b1,….,bm are the m distinct elements of S2, then S1 x S2 = Uin =1 (ai x S2).

For if x is an arbitrary element of S1 x S2 , then x = (a, b) where a Î S1 and b Î S2. Thus, a = ai for some i and b = bj for some j. Thus, x = (ai, bj) Î(ai x S2) and therefore x Î Uni =1(ai x S2).

Conversely, if x Î Uin =1(ai x S2), then x Î (ai x S2) for some i and thus x = (ai, bj) where bj is some element of S2. Therefore, x Î S1 x S2.

Next observe that (ai x S2) and (aj x S2) are disjoint if i ≠ j since if

x Î (ai x S2) ∩ (aj x S2) then x = (ai, bk) for some k and x = (aj, b1) for some l.

But then (ai, bk) = (aj, bl) implies that ai = aj and bk = bl. But since i ≠ j , ai ≠ a j.

Thus, we conclude that S1 x S2 is the disjoint union of the sets (ai x S2). Furthermore |ai x S2| = |S2| since there is obviously a one-to-one correspondence between the sets ai x S2 and

S2, namely, (ai, bj) → bj.

Then by the sum rule |S1 x S2| = ∑nni=1 | ai x S2|

7.(n summands) |S2| + |S2| +….... + |S2|

8.n |S2|

9.nm.

Therefore, we have proven the product rule for two sets. The general rule follows by mathematical induction.

We can reformulate the product rule in terms of events. If events E1, E2 , …., En can happen e1, e2,…., and en ways, respectively, then the sequence of events E1 first, followed by E2,…., followed by En can happen e1e2 …en ways.

In terms of choices, the product rule is stated thus: If a first object can be chosen e1 ways, **a second e2 ways , …, and an nth object can be made in e1e2….en ways.**

Combinations & Permutations:

Definition.

A combination of n objects taken r at a time (called an r-combination of n objects) is an unordered selection of r of the objects.

A permutation of n objects taken r at a time (also called an r-permutation of n objects) is an ordered selection or arrangement of r of the objects.

Note that we are simply defining the terms r-combinations and r-permutations here and have not mentioned anything about the properties of the n objects.

For example, these definitions say nothing about whether or not a given element may appear more than once in the list of n objects.

In other words, it may be that the n objects do not constitute a set in the normal usage of the word.

SOLVED PROBLEMS

Example1. Suppose that the 5 objects from which selections are to be made are: a, a, a, b, c. then the 3-combinations of these 5 objects are : aaa, aab, aac, abc. The permutations are:

aaa, aab, aba, baa, aac, aca, caa,
abc, acb, bac, bca, cab, cba.

Neither do these definitions say anything about any rules governing the selection of the r-objects: on one extreme, objects could be chosen where all repetition is forbidden, or on the other extreme, each object may be chosen up to t times, or then again may be some rule of selection between these extremes; for instance, the rule that would allow a given object to be repeated up to a certain specified number of times.

We will use expressions like {3 . a , 2. b ,5.c} to indicate either

(1) that we have 3 + 2 + 5 =10 objects including 3a's , 2b's and 5c's, or (2) that we have 3 objects a, b, c, where selections are constrained by the conditions that a can be selected at most three times, b can be selected at most twice, and c can be chosen up to five times.

The numbers 3, 2 and 5 in this example will be called repetition numbers.

Example 2 The 3-combinations of {3. a, 2. b, 5. c} are:

aaa, aab, aac, abb,
abc, ccc, ccb, cca,
cbb.

Example 3**. The 3-combinations of {3 . a, 2. b, 2. c , 1. d} are:**

aaa, aab, aac, aad, bba, bbc, bbd,
cca, ccb, ccd, abc, abd, acd, bcd.

In order to include the case where there is no limit on the number of times an object can be repeated in a selection (except that imposed by the size of the selection) we use the **symbol** ∞ **as a repetition number to mean that an object can occur an infinite number of times.**

Example 4. The 3-**combinations of {∞. a, 2.b, ∞.c} are the same as in Example 2 even though** a and c can be repeated an infinite number of times. This is because, in 3-combinations, 3 is the limit on the number of objects to be chosen.

If we are consideri**ng selections where each object has ∞ as its repetition number then** we designate such selections as selections with unlimited repetitions. In particular, a selection of r objects in this case will be called r-combinations with unlimited repetitions and any ordered arrangement of these r objects will be an r-permutation with unlimited repetitions.

Example5 **The combinations of a ,b, c, d with unlimited repetitions are the 3-combinations of {∞ . a , ∞. b, ∞. c, ∞. d}. These are 20 such 3**-combinations, namely:

aaa, aab, aac, aad,
bbb, bba, bbc, bbd,
ccc, cca, ccb, ccd,
ddd, dda, ddb, ddc,
abc, abd, acd, bcd.

2-combinations with Unlimited Repetitions	**2-permutations with Unlimited Repetitions**
aa	Aa
ab	ab, ba
ac	ac, ca
ad	ad, da
bb	Bb
bc	bc, cb
bd	bd, db
cc	Cc
cd	cd, dc
dd	Dd
10	16

Moreover, there are 43 = 64 of 3-permutations with unlimited repetitions since the first position can be filled 4 ways (with a, b, c, or d), the second position can be filled 4 ways, and likewise for the third position.

The 2-**permutations of {∞. a, ∞. b, ∞. c, ∞. d} do not present such a formi**dable list and so we tabulate them in the following table.

We list some more examples just for concreteness. We might, for example, consider selections of {∞.a, ∞. b, ∞. c} where b can be chosen only even number of times. Thus, 5-combinations with these repetition numbers and this constraint would be those 5-combinations with unlimited repetitions and where b is chosen 0, 2, or 4 times.

Example6 The 3-**combinations of {∞ .a, ∞ .b, 1 .c,1 .d} where b can be chosen only an even** number of times are the 3-combinations of a, b, c, d where a can be chosen up 3 times, b can be chosen 0 or 2 times, and c and d can be chosen at most once. The 3-cimbinations subject to these constraints are:

aaa, aac, aad, bbc, bbd, acd.

As another example, we might be interested in, selections of {∞.a, 3.b, 1.c} where a can **be chosen a prime number of times. Thus, the 8-combinations subject to these constraints would be all those 8-combinations where a can be chosen 2, 3, 5, or 7 times, b can chosen up to 3 times, and c can be chosen at most once.**

There are, as we have said, an infinite variety of constraints one could place on selections. You can just let your imagination go free in conjuring up different constraints on the selection, would constitute an r-combination according to our definition. Moreover, any arrangement of these r objects would constitute an r-permutation.

While there may be an infinite variety of constraints, we are primarily interested in two major types: one we have already described—combinations and permutations with unlimited repetitions, the other we now describe.

If the repetition numbers are all 1, then selections of r objects are called r-combinations without repetitions and arrangements of the r objects are r-permutations without repetitions. We remind you that r-combinations without repetitions are just subsets of the n elements containing exactly r elements. Moreover, we shall often drop the repetition number 1 when considering r-combinations without repetitions. For example, when considering r-combinations of {a, b, c, d} we will mean that each repetition number is 1 unless otherwise designated, and, of course, we mean that in a given selection an element need not be chosen at all, but, if it is chosen, then in this selection this element cannot be chosen again.

Example7. **Suppose selections are to be made from the four objects a, b, c, d.**

2-combinations without Repetitions	2-Permutations without Repetitions
ab	**ab, ba**
ac	ac, ca
ad	ad, da
bc	bc, cb

bd	bd, db
cd	cd, dc
6	12

There are six 2-combinations without repetitions and to each there are two 2-permutations giving a total of twelve 2-permutations without repetitions.

Note that total number of 2-combinations with unlimited repetitions in Example 5 included six 2-combinations without repetitions of Example.7 and as well 4 other 2-combinations where repetitions actually occur. Likewise, the sixteen 2-permutations with unlimited repetitions included the twelve 2-permutations without repetitions.

3-combinations without Repetitions	3-Permutations without Repetitions
abc	abc, acb, bac, bca, cab, cba
abd	abd, adb, bad, bda, dab, dba
acd	acd, adc, cad, cda, dac, dca
bcd	bcd, bdc, cbd, cdb, dbc, dcb
4	24

Note that to each of the 3-combinations without repetitions there are 6 possible 3- permutations without repetitions. Momentarily, we will show that this observation can be generalized.

Combinations And Permutations With Repetitions:

General formulas for enumerating combinations and permutations will now be presented. At this time, we will only list formulas for combinations and permutations without repetitions or with unlimited repetitions. We will wait until later to use generating functions to give general techniques for enumerating combinations where other rules govern the selections.

Let P (n, r) denote the number of r-permutations of n elements without repetitions.

Theorem 5.3.1.(Enumerating r-permutations without repetitions).

$$P(n, r) = n(n-1)\ldots\ldots (n - r + 1) = n! / (n-r)!$$

Proof. Since there are n distinct objects, the first position of an r-permutation may be filled in n ways. This done, the second position can be filled in n-1 ways since no repetitions are allowed and there are n – 1 objects left to choose from. The third can be filled in n-2 ways. By applying the product rule, we conduct that

$$P (n, r) = n(n-1)(n-2)\ldots\ldots (n - r + 1).$$

From the definition of factorials, it follows that

$$P (n, r) = n! / (n-r)!$$

When r = n, this formula becomes

$$P (n, n) = n! / 0! = n!$$

When we explicit reference to r is not made, we assume that all the objects are to be arranged; thus we talk about the permutations of n objects we mean the case r=n. Corollary 1. There are n! permutations of n distinct objects.

Example 1.

There are 3! = 6 permutations of {a, b, c}.

There are 4! = 24 permutations of (a, b, c, d). The number of 2-permutations {a, b, c, d, e} is P(5, 2) = 5! /(5 - 2)! = 5 x 4 = 20.

The number of 5-letter words using the letters a, b, c, d, and e at most once is P (5, 5) = 120.

Example 2 There are P (10, 4) = 5,040 4-digit numbers that contain no repeatd digits since each such number is just an arrangement of four of the digits 0, 1, 2, 3 , …., 9 (leading zeroes are allowed). There are P (26, 3) P(10, 4) license plates formed by 3 distinct letters followed by 4 distinct digits.

Example3. **In how many ways can 7 women and 3 men be arranged in a row if the 3 men must always stand next to each other?**

There are 3! ways of arranging the 3 men. Since the 3 men always stand next to each **other, we treat them as a single entity, which we denote by X. Then if W1, W2, ….., W7** represents the women, we next are interested in the number of ways of arranging {X, W1, W2, **W3,……., W7}. There are 8! permutations these 8 objects. Hence there are (3!)** (8!) permutations altogether. (of course, if there has to be a prescribed order of an arrangement on the

3 men then there are only 8! total permutations).

Example4. In how many ways can the letters of the English alphabet be arranged so that there

are exactly 5 letters between the letters a and b?

There are P (24, 5) ways to arrange the 5 letters between a and b, 2 ways to place a and b, and then 20! ways to arrange any 7-letter word treated as one unit along with the remaining 19 letters. The total is P (24, 5) (20!) (2).

permutations **for the objects are being arranged in a line. If instead of arranging objects in a line, we arrange them in a circle, then the number of permutations decreases.**

Example 5. **In how many ways can 5 children arrange themselves in a ring?**

Solution. Here, the 5 children are not assigned to particular places but are only arranged relative to one another. Thus, the arrangements (see Figure 2-3) are considered the same if the children are in the same order clockwise. Hence, the position of child C1 is immaterial and it is only the position of the 4 other children relative to C1 that counts. Therefore, keeping C1 fixed in position, there are 4! arrangements of the remaining children.

Binomial Coefficients: In mathematics, the **binomial coefficient** $\binom{n}{k}$ is the coefficient of the x^k **term in the polynomial expansion of the binomial power** $(1 + x)^n$ **.**

In combinatorics, $\binom{n}{k}$ is interpreted as the number of *k*-element subsets (the *k*-combinations) of an *n*-element set, that is the number of ways that *k* things can be "chosen" from a set of *n* things. Hence, $\binom{n}{k}$ is often read as "*n* choose *k*" and is called the choose function of *n* and *k*. The notation $\binom{n}{k}$ was introduced by Andreas von Ettingshausen in 182, although the numbers were already known centuries before that (see Pascal's triangle). Alternative notations include C(*n*,*k*), $_nC_k$, ^nC_k, C^n_k, C^k_n, in all of which the C stands for combinations or *choices*.

For natural numbers (taken to include 0) *n* and *k*, the binomial coefficient $\binom{n}{k}$ can be definedas the coefficient of the monomial X^k in the expansion of $(1 + X)^n$. The same coefficient also occurs (if $k \leq n$) in the binomial formula

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

(valid for any elements x,y of a commutative ring), which explains the name "binomial coefficient".

Another occurrence of this number is in combinatorics, where it gives the number of ways, disregarding order, that a k objects can be chosen from among n objects; more formally, the number of k-element subsets (or k-combinations) of an n-element set. This number can be seen to be equal to the one of the first definition, independently of any of the formulas below to compute it: if in each of the n factors of the power $(1 + X)^n$ one temporarily labels the term X with an index i (running from 1 to n), then each subset of k indices gives after expansion a contribution X^k, and the coefficient of that monomial in the result will be the number of such subsets. This shows in particular that $\binom{n}{k}$ is a natural number for any natural numbers n and k. There are many other combinatorial interpretations of binomial coefficients (counting problems for which the answer is given by a binomial coefficient expression), for instance the number of words formed of n bits (digits 0 or 1) whose sum is k, but most of these are easily seen to be equivalent to counting k-combinations.

Several methods exist to compute the value of $\binom{n}{k}$ without actually expanding a binomial power or counting k-combinations.

Binomial Multinomial theorems:

Binomial theorem:

In elementary algebra, the binomial theorem describes the algebraic expansion of powers of a binomial. According to the theorem, it is possible to expand the power $(x + y)^n$ into a sum involving terms of the form $ax^b y^c$, where the coefficient of each term is a positive integer, and the sum of the exponents of x and y in each term is n. For example,

$$(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4.$$

The coefficients appearing in the binomial expansion are known as binomial coefficients. They are the same as the entries of Pascal's triangle, and can be determined by a simple formula involving factorials. These numbers also arise in combinatorics, where the coefficient of $x^{n-k} y^k$ is equal to the number of different combinations of k elements that can be chosen from an n-element set.

According to the theorem, it is possible to expand any power of *x* + *y* into a sum of the form

$$(x+y)^n = \binom{n}{0}x^n y^0 + \binom{n}{1}x^{n-1}y^1 + \binom{n}{2}x^{n-2}y^2 + \binom{n}{3}x^{n-3}y^3 + \cdots$$
$$\cdots + \binom{n}{n-1}x^1 y^{n-1} + \binom{n}{n}x^0 y^n,$$

where $\binom{n}{k}$ denotes the corresponding binomial coefficient. Using summation notation, the formula above can be written

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k.$$

This formula is sometimes referred to as the **Binomial Formula** or the **Binomial Identity**.

A variant of the binomial formula is obtained by substituting 1 for *x* and *x* for *y*, so that it involves only a single variable. In this form, the formula reads

$$(1+x)^n = \binom{n}{0}x^0 + \binom{n}{1}x^1 + \binom{n}{2}x^2 + \cdots + \binom{n}{n-1}x^{n-1} + \binom{n}{n}x^n,$$

or equivalently

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k.$$

Multinomial theorem:

In mathematics, the **multinomial theorem** says how to write a power of a sum in terms of powers of the terms in that sum. It is the generalization of the binomial theorem to polynomials.

For any positive integer *m* and any nonnegative integer *n*, the multinomial formula tells us how a polynomial expands when raised to an arbitrary power:

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{k_1+k_2+\cdots+k_m=n} \binom{n}{k_1, k_2, \ldots, k_m} x_1^{k_1} x_2^{k_2} \cdots x_m^{k_m}.$$

The summation is taken over all sequences of nonnegative integer indices k_1 through k_m such the sum of all k_i is n. That is, for each term in the expansion, the exponents must add up to n.

Also, as with the binomial theorem, quantities of the form x^0 that appear are taken to equal 1 (even when x equals zero). Alternatively, this can be written concisely using multiindices as

$$(x_1 + \cdots + x_m)^n = \sum_{|\alpha|=n} \binom{n}{\alpha} x^\alpha$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ and $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_m^{\alpha_m}$.

Example

$(a + b + c)^3 = a^3 + b^3 + c^3 + 3a^2b + 3a^2c + 3b^2a + 3b^2c + 3c^2a + 3c^2b + 6abc.$

We could have calculated each coefficient by first expanding

$(a + b + c)^2 = a^2 + b^2 + c^2 + 2ab + 2bc + 2ac$, then self-multiplying it again to get $(a + b + c)^3$

(and then if we were raising it to higher powers, we'd multiply it by itself even some more).

However this process is slow, and can be avoided by using the multinomial theorem. The multinomial theorem "solves" this process by giving us the closed form for any coefficient we might want. It is possible to "read off" the multinomial coefficients from the terms by using the multinomial coefficient formula. For example:

$a^2 b^0 c^1$ has the coefficient $\binom{3}{2,0,1} = \frac{3!}{2!\cdot 0!\cdot 1!} = \frac{6}{2\cdot 1\cdot 1} = 3$

$a^1 b^1 c^1$ has the coefficient $\binom{3}{1,1,1} = \frac{3!}{1!\cdot 1!\cdot 1!} = \frac{6}{1\cdot 1\cdot 1} = 6$.

We could have also had a 'd' variable, or even more variables—hence the *multinomial* theorem.

The principles of Inclusion – Exclusion:

Let $|A|$ denote the cardinality of set A, then it follows immediately that

$$|A \bigcup B| = |A| + |B| - |A \bigcap B|, \qquad (1)$$

where $\bigcup$ denotes union, and $\bigcap$ denotes intersection. The more general statement

$$\left|\bigcup_{i=1}^{N} E_i\right| \leq \sum_{i=1}^{N} |E_i|, \qquad \textbf{(2)}$$

also holds, and is known as Boole's inequality.

This formula can be generalized in the following beautiful manner. Let $\mathcal{A} = \{A_i\}_{i=1}^{p}$ bea *p*-system of consisting of sets , ..., A_p, then

$$|A_1 \cup A_2 \cup \ldots \cup A_p| = \sum_{1\leq i\leq p} |A_i| - \sum_{1\leq i_1 < i_2 \leq p} |A_{i_1} \cap A_{i_2}| + \sum_{1\leq i_1 < i_2 < i_3 \leq p} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| - \ldots + (-1)^{p-1} |A_1 \cap A_2^3 \cap \cdots \cap A_p$$

where the sums are taken over *k*-subsets of $\mathcal{A}$. This formula holds for infinite sets S as well as finite sets.
The principle of inclusion-exclusion was used by Nicholas Bernoulli to solve the recontres problem of finding the number of derangements.

For example, for the three subsets $A_1 = \{2, 3, 7, 9, 10\}$, $A_2 = \{1, 2, 3, 9\}$, and $A_3 = \{2, 4, 9, 10\}$ of $S = \{1, 2, \ldots, 10\}$, the following table summarizes the termsappearing the sum.

#	term	set	length
1	A_1	{2, 3, 7, 9, 10	5
	A_2	{1, 2, 3, 9	4
	A_3	{2, 4, 9, 10	4
2	$A_1 \cap A_2$	{2, 3, 9	3
	$A_1 \cap A_3$	{2, 9, 10	3
	$A_2 \cap A_3$	{2, 9	2
3		2, 9	2

$$(5+4+4)-(3+3+2)+2=7$$

$$A_1 \cap A_2 \cap A_3 \quad \{ \quad \}$$

$|A_1 \cup A_2 \cup A_3|$ **is therefore equal to, corresponding to the seven elements** $A_1 \cup A_2 \cup A_3 = \{1, 2, 3, 4, 7, 9, 10\}$**.**

Pigeon hole principles and its application:

The statement of the *Pigeonhole Principle*:

If m pigeons are put into m pigeonholes, there is an empty hole *iff* there's a hole with more than one pigeon.
If n > m pigeons are put into m pigeonholes, there's a hole with more than one pigeon.

Example:
Consider a chess board with two of the diagonally opposite corners removed. Is it possible to cover the board with pieces of domino whose size is exactly two board squares?

Solution

No, it's not possible. Two diagonally opposite squares on a chess board are of the same color. Therefore, when these are removed, the number of squares of one color exceeds by 2 the number of squares of another color. However, every piece of domino covers exactly two squares and these are of different colors. Every placement of domino pieces establishes a 1-1 correspondence between the set of white squares and the set of black squares. If the two sets have different number of elements, then, by the Pigeonhole Principle, no 1-1 correspondence between the two sets is possible.

Generalizations of the pigeonhole principle

A generalized version of this principle states that, if n discrete objects are to be allocated to m containers, then at least one container must hold no fewer than $\lceil n/m \rceil$ objects, where $\lceil x \rceil$ is the ceiling function, denoting the smallest integer larger than or equal to x. Similarly, at least one container must hold no more than $\lfloor n/m \rfloor$ objects, where $\lfloor x \rfloor$ is the floor function, denoting the largest integer smaller than or equal to x.

A probabilistic generalization of the pigeonhole principle states that if n pigeons are randomly put into m pigeonholes with uniform probability $1/m$, then at least one pigeonhole will hold more than one pigeon with probability

$$1 - \frac{(m)_n}{m^n},$$

where $(m)_n$ is the falling factorial $m(m - 1)(m - 2)...(m - n + 1)$. For $n = 0$ and for $n = 1$ (and $m > 0$), that probability is zero; in other words, if there is just one pigeon, there cannot be a conflict. For $n > m$ (more pigeons than pigeonholes) it is one, in which case it coincides with the ordinary pigeonhole principle. But even if the number of pigeons does not exceed the number of pigeonholes ($n \leq m$), due to the random nature of the assignment of pigeons to pigeonholes there is often a substantial chance that clashes will occur. For example, if 2 pigeons are randomly assigned to 4 pigeonholes, there is a 25% chance that at least one pigeonhole will hold more than one pigeon; for 5 pigeons and 10 holes, that probability is 69.76%; and for 10 pigeons and 20 holes it is about 93.45%. If the number of holes stays fixed, there is always a greater probability of a pair when you add more pigeons. This problem is treated at much greater length at birthday paradox.

A further probabilistic generalisation is that when a real-valued random variable X has a finite mean $E(X)$, then the probability is nonzero that X is greater than or equal to $E(X)$, and similarly the probability is nonzero that X is less than or equal to $E(X)$. To see that this implies the standard pigeonhole principle, take any fixed arrangement of n pigeons into m holes and let X be the number of pigeons in a hole chosen uniformly at random. The mean of X is n/m, so if there are more pigeons than holes the mean is greater than one. Therefore, X is sometimes at least 2.

Applications:

The pigeonhole principle arises in computer science. For example, collisions are inevitable in a hash table because the number of possible keys exceeds the number of indices in the array. No hashing algorithm, no matter how clever, can avoid these collisions. This principle also proves that any general-purpose lossless compression algorithm that makes at least one input file smaller will make some other input file larger. (Otherwise, two files would be compressed to the same smaller file and restoring them would be ambiguous.)

A notable problem in mathematical analysis is, for a fixed irrational number *a*, to show that the set {[*na*]: *n* is an integer} of fractional parts is dense in [0, 1]. After a moment's thought, one finds that it is not easy to explicitly find in tegers , such that | − | < , where > 0 is a *n m* C *name* small positive number and a is some arbitrary irrational number. But if one takes M such that

$1/M < e$, by the pigeonhole principle there must be n_1, n_2 {1, 2, ..., M + 1} such that n_1a and n_2a are in the same integer subdivision of size $1/M$ (there are only M such subdivisions between consecutive

integers). In particular, we can find n_1, n_2 such that n_1a is in $(p + k/M, p + (k + 1)/M)$, and n_2a is in $(q + k/M, q + (k + 1)/M)$, for some p, q integers and k in $\{0, 1, ..., M - 1\}$. We can then easily verify that $(n_2 - n_1)a$ is in $(q - p - 1/M, q - p + 1/M)$. This implies that $[na] < 1/M < e$, where $n = n_2 - n_1$ or $n = n_1 - n_2$. This shows that 0 is a limit point of $\{[na]\}$. We can then use this fact to prove the case for p in (0, 1]: find n such that $[na] < 1/M < e$; then if p (0, $1/M$], we are done. Otherwise in (/ , (+ 1)/], and by setting = sup{ : [ ] < /

C

$|[(k + 1)na] - p| < 1/M < e$.