Sirindhorn International Institute of Technology
Thammasat University
School of Information, Computer and Communication Technology

IES302 2011/1 Part I.1 Dr.Prapun

1 Probability and You

Whether you like it or not, probabilities rule your life. If you have ever tried to make a living as a gambler, you are painfully aware of this, but even those of us with more mundane life stories are constantly affected by these little numbers.

Example 1.1. Some examples from daily life where probability calculations are involved are the determination of insurance premiums, the introduction of new medications on the market, opinion polls, weather forecasts, and DNA evidence in courts. Probabilities also rule who you are. Did daddy pass you the X or the Y chromosome? Did you inherit grandma's big nose?

Meanwhile, in everyday life, many of us use probabilities in our language and say things like "I'm 99% certain" or "There is a one-in-a-million chance" or, when something unusual happens, ask the rhetorical question "What are the odds?". [15, p 1]

1.1 Randomness

1.2. Many clever people have thought about and debated what randomness really is, and we could get into a long philosophical discussion that could fill up a whole book. Let's not. The French mathematician Laplace (1749–1827) put it nicely:

> "Probability is composed partly of our ignorance, partly of our knowledge."

Inspired by Laplace, let us agree that you can use probabilities whenever you are faced with uncertainty. [15, p 2]

1.3. Random phenomena arise because of [10]:

(a) our partial ignorance of the generating mechanism

(b) the laws governing the phenomena may be fundamentally random (as in quantum mechanics)

(c) our unwillingness to carry out exact analysis because it is not worth the trouble

Example 1.4. Communication Systems [22]: The essence of communication is randomness.

(a) **Random Source**: The transmitter is connected to a random source, the output of which the receiver cannot predict with certainty.

- If a listener knew in advance exactly what a speaker would say, and with what intonation he would say it, there would be no need to listen!

(b) **Noise**: There is no communication problem unless the transmitted signal is disturbed during propagation or reception in a random way.

(c) Probability theory is used to *evaluate the performance* of communication systems.

Example 1.5. Random numbers are used directly in the transmission and security of data over the airwaves or along the Internet.

(a) A radio transmitter and receiver could switch transmission frequencies from moment to moment, seemingly at random, but nevertheless in synchrony with each other.

(b) The Internet data could be credit-card information for a consumer purchase, or a stock or banking transaction secured by the clever application of random numbers.

Example 1.6. Randomness is an essential ingredient in games of all sorts, computer or otherwise, to make for unexpected action and keen interest.

Example 1.7. On a more profound level, quantum physicists teach us that everything is governed by the laws of probability. They toss around terms like the Schrödinger wave equation and Heisenberg's uncertainy principle, which are much too difficult for most of us to understand, but one thing they do mean is that the fundamental laws of physics can only be stated in terms of probabilities. And the fact that Newton's deterministic laws of physics are still useful can also be attributed to results from the theory of probabilities. [15, p 2]

1.8. Most people have preconceived notions of randomness that often differ substantially from true randomness. Truly random data sets often have unexpected properties that go against intuitive thinking. These properties can be used to test whether data sets have been tampered with when suspicion arises. [20, p 191]

- [11, p 174]: "people have a very poor conception of randomness; they do not recognize it when they see it and they cannot produce it when they try"

Example 1.9. Apple ran into an issue with the random shuffling method it initially employed in its iPod music players: true randomness sometimes produces repetition, but when users heard the same song or songs by the same artist played back-to-back, they believed the shuffling wasn't random. And so the company made the feature "less random to make it feel more random," said Apple founder Steve Jobs. [11, p 175]

1.2 Background on some frequently used examples

Probabilists love to play with coins and dice. We like the idea of tossing coins, rolling dice, and drawing cards as experiments that have equally likely outcomes.

1.10. ***Coin flipping*** or ***coin tossing*** is the practice of throwing a coin in the air to observe the outcome.

When a **coin** is tossed, it does not necessarily fall heads or tails; it can roll away or stand on its edge. Nevertheless, we shall agree to regard "**head**" (**H**) and "**tail**" (**T**) as the only possible outcomes of the experiment. [3, p 7]

- Typical experiment includes
  - "Flip a coin N times. Observe the sequence of heads and tails" or "Observe the number of heads."

1.11. Historically, ***dice*** is the plural of ***die***, but in modern standard English dice is used as both the singular and the plural. [Excerpted from Compact Oxford English Dictionary.]

- Usually assume six-sided dice
- Usually observe the number of dots on the side facing upwards.

1.12. A complete set of **cards** is called a pack or **deck**.

(a) The subset of cards held at one time by a player during a game is commonly called a **hand**.

(b) For most games, the cards are assembled into a deck, and their order is randomized by **shuffling**.

(c) A standard deck of 52 cards in use today includes thirteen ranks of each of the four French suits.

- The four suits are called spades ($\spadesuit$), clubs ($\clubsuit$), hearts ($\heartsuit$), and diamonds ($\diamondsuit$). The last two are red, the first two black.

(d) There are thirteen face values ($2, 3, \ldots, 10$, jack, queen, king, ace) in each suit.

- Cards of the same face value are called of the same **kind**.
- "court" or face card: a king, queen, or jack of any suit.

(e) For our purposes, playing bridge means distributing the cards to four players so that each receives thirteen cards. Playing poker, by definition, means selecting five cards out of the pack.

1.3 A Glimpse at Probability

1.13. Probabilities are used in situations that involve ***randomness***. A ***probability*** is a number used to describe how likely something is to occur, and ***probability*** (without indefinite article) is the study of probabilities. It is the art of ***being certain of how uncertain you are***. [15, p 2–4] If an event is certain to happen, it is given a probability of 1. If it is certain not to happen, it has a probability of 0. [5, p 66]

1.14. Probabilities can be expressed as fractions, as decimal numbers, or as percentages. If you toss a coin, the probability to get heads is 1/2, which is the same as 0.5, which is the same as 50%. There are no explicit rules for when to use which notation.

- In daily language, proper fractions are often used and often expressed, for example, as "one in ten" instead of 1/10 ("one tenth"). This is also natural when you deal with equally likely outcomes.
- **Decimal numbers** are more common in technical and scientific reporting when probabilities are calculated from data. Percentages are also common in daily language and often with "chance" replacing "probability."
- Meteorologists, for example, typically say things like "there is a 20% chance of rain." The phrase "the probability of rain is 0.2" means the same thing.
- When we deal with probabilities from a theoretical viewpoint, we always think of them as numbers between 0 and 1, not as percentages.
- See also 3.7.

[15, p 10]

Definition 1.15. Important terms [10]:

(a) An activity or procedure or observation is called a **random experiment** if its outcome cannot be predicted precisely because the conditions under which it is performed cannot be predetermined with sufficient accuracy and completeness.

- The term "experiment" is to be construed loosely. We do not intend a laboratory situation with beakers and test tubes.
- Tossing/flipping a coin, rolling a die, and drawing a card from a deck are some examples of random experiments.

(b) A random experiment may have several separately identifiable **outcomes**. We define the **sample space** Ω as a collection of all possible (separately identifiable) outcomes/results/measurements of a random experiment. Each outcome (ω) is an element, or sample point, of this space.

- Rolling a dice has six possible identifiable outcomes ($1, 2, 3, 4, 5$, and 6).

(c) **Events** are sets (or classes) of outcomes meeting some specifications.

- Any event is a subset of Ω.
- Intuitively, an event is a statement about the outcome(s) of an experiment.
- For our class, it may be less confusing to allow event A to be any collection of outcomes (, i.e. any subset of Ω).
  - In more advanced courses, when we deal with uncountable Ω, we limit our interest to only some subsets of Ω. Technically, the collection of these subsets must form a σ-algebra.

The goal of probability theory is to compute the probability of various events of interest. Hence, we are talking about a set function which is defined on (some class of) subsets of Ω.

Example 1.16. The statement "when a coin is tossed, the probability to get heads is l/2 (50%)" is a *precise* statement.

(a) It tells you that you are as likely to get heads as you are to get tails.

(b) Another way to think about probabilities is in terms of **average long-term behavior**. In this case, if you toss the coin repeatedly, in the long run you will get *roughly* 50% heads and 50% tails.

[15, p 4]

1.17. *Long-run frequency interpretation*: If the probability of an event A in some actual physical experiment is p, then we believe that if the experiment is *repeated independently* over and over again, then a theorem called the ***law of large numbers*** (LLN) states that, in the long run, the event A will happen approximately $100p\%$ of the time. In other words, if we repeat an experiment a large number of times then the fraction of times the event A occurs will be close to $P(A)$.

Definition 1.18. Let A be one of the events of a random experiment. If we conduct a sequence of n independent trials of this experiment, and if the event A occurs in $N(A, n)$ out of these n trials, then the fraction

is called the **relative frequency** of the event A in these n trials.

The long-run frequency interpretation mentioned in 1.17 can be restated as

$$P(A) \text{ “=” } \lim_{n\to\infty} \frac{N(A,n)}{n}.$$

1.19. In terms of practical range, probability theory is comparable with ***geometry***; both are branches of applied mathematics that are directly linked with the problems of daily life. But while pretty much anyone can call up a natural feel for geometry to some extent, many people clearly have trouble with the development of a good intuition for probability.

- Probability and intuition do not always agree. ***In no other branch of mathematics is it so easy to make mistakes as in probability theory.***
- Students facing difficulties in grasping the concepts of probability theory might find comfort in the idea that even the genius Leibniz, the inventor of differential and integral calculus along with Newton, had difficulties in calculating the probability of throwing 11 with one throw of two dice.

[20, p 4]

2 Review of Set Theory

2.1. If ω is a member of a set A, we write $\omega \in A$.

Definition 2.2. Basic set operations (set algebra)

- Complementation: $A^c = \{\omega : \omega \notin A\}$.
- Union: $A \cup B = \{\omega : \omega \in A \text{ or } \omega \in B\}$
  - Here "or"is inclusive; i.e., if $\omega \in A$, we permit ω to belong either to A or to B or to both.
- Intersection: $A \cap B = \{\omega : \omega \in A \text{ and } \omega \in B\}$
  - Hence, $\omega \in A$ if and only if ω belongs to both A and B.
  - $A \cap B$ is sometimes written simply as AB.
- The ***set difference*** operation is defined by $B \setminus A = B \cap A^c$.
  - $B \setminus A$ is the set of $\omega \in B$ that do not belong to A.
  - When $A \subset B$, $B \setminus A$ is called the complement of A in B.

2.3. Basic Set Identities:

- Idempotence: $(\mathrm{A}^c)^c = \mathrm{A}$
- Commutativity (symmetry):

$$A \cup B = B \cup A\ ,\ A \cap B = B \cap A$$

- Associativity:
  - $A \cap (B \cap C) = (A \cap B) \cap C$
  - $A \cup (B \cup C) = (A \cup B) \cup C$
- Distributivity

- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

- **de Morgan laws**
  - $(A \cup B)^c = A^c \cap B^c$
  - $(A \cap B)^c = A^c \cup B^c$

2.4. Venn diagram is very useful in set theory. Many identities can be read out simply by examining Venn diagrams.

2.5. Disjoint Sets:

- Sets A and B are said to be ***disjoint*** $(A \perp B)$ if and only if $A \cap B = \emptyset$. (They do not share member(s).)
- A collection of sets $(A_i : i \in I)$ is said to be ***pairwise disjoint*** or mutually exclusive [7, p. 9] if and only if $\mathrm{A}_i \cap \mathrm{A}_j = \emptyset$ when $i \neq j$.

2.6. For a set of sets, to avoid the repeated use of the word "set", we will call it a **collection/class/family** of sets.

Definition 2.7. Given a set S, a collection $\Pi = (A_\alpha : \alpha \in I)$ of subsets[1] of S is said to be a ***partition*** of S if

(a) $S = \bigcup A_{\alpha \in I}$ and

(b) For all $i \neq j$, $A_i \perp A_j$ (pairwise disjoint).

Remarks:

- The subsets A_α, $\alpha \in I$ are called the **parts** of the partition.
- A part of a partition may be empty, but usually there is no advantage in considering partitions with one or more empty parts.

Example 2.8 (Slide:maps)**.**

[1] In this case, the subsets are indexed or labeled by α taking values in an index or label set I

Example 2.9. Let E be the set of students taking IES302

Definition 2.10. The ***cardinality*** (or size) of a collection or set A, denoted $|A|$, is the number of elements of the collection. This number may be finite or infinite.

- An infinite set A is said to be ***countable*** if the elements of A can be enumerated or listed in a sequence: $a_1, a_2, \ldots$.
    - Empty set and finite sets are also said to be countable.
- By a ***countably infinite*** set, we mean a countable set that is not finite. Examples of such sets include
    - the set $\mathbb{N} = \{1, 2, 3, \ldots\}$ of natural numbers,
    - the set $\{2k : k \in \mathbb{N}\}$ of all even numbers,
    - the set $\{2k + 1 : k \in \mathbb{N}\}$ of all odd numbers,
    - the set $\mathbb{Z}$ of integers,
    - the set $\mathbb{Q}$ of all rational numbers,
    - the set $\mathbb{Q}^+$ of positive rational numbers,
    - the set of all finite-length sequences of natural numbers,
    - the set of all finite subsets of the natural numbers.
- A ***singleton*** is a set with exactly one element.
    - Ex. $\{1.5\}$, $\{.8\}$, $\{\pi\}$.

Set Theory	Probability Theory
Set	Event
Universal set	Sample Space (Ω)
Element	Outcome (ω)

Table 1: The terminology of set theory and probability theory

- *Caution*: Be sure you understand the difference between the outcome -8 and the event $\{-8\}$, which is the set consisting of the single outcome -8.

2.11. We can categorize sets according to their cardinality:

Example 2.12. Example of uncountable sets[2]:

- $\mathbb{R} = (-\infty, \infty)$
- interval $[0, 1]$
- interval $(0, 1]$
- $(2, 3) \cup [5, 7)$

Definition 2.13. Probability theory renames some of the terminology in set theory. See Table 1 and Table 2.

- Sometimes, ω's are called states, and Ω is called the state space.

[2]We use a technique called diagonal argument to prove that a set is not countable and hence uncountable.

	Event Language
A	A occurs
A^c	A does not occur
$A \cup B$	Either A or B occur
$A \cap B$	Both A and B occur

Table 2: Event Language

2.14. Because of the mathematics required to determine probabilities, probabilistic methods are divided into two distinct types, discrete and continuous. A discrete approach is used when the number of experimental outcomes is finite (or infinite but countable). A continuous approach is used when the outcomes are continuous (and therefore infinite). It will be important to keep in mind which case is under consideration since otherwise, certain paradoxes may result.

3 Classical Probability

Classical probability, which is based upon the ratio of the number of outcomes favorable to the occurrence of the event of interest to the total number of possible outcomes, provided most of the probability models used prior to the 20th century. It is the first type of probability problems studied by mathematicians, most notably, Frenchmen Fermat and Pascal whose 17th century correspondence with each other is usually considered to have started the systematic study of probabilities. [15, p 3] Classical probability remains of importance today and provides the most accessible introduction to the more general theory of probability.

Definition 3.1. Given a finite sample space Ω, the ***classical probability*** of an event A is

$$P(A) = \frac{|A|}{|\Omega|} \tag{1}$$

[4, Defn. 2.2.1 p 58]. In traditional language, a probability is a fraction in which the bottom represents the number of possible outcomes, while the number on top represents the number of outcomes in which the event of interest occurs.

- Assumptions: When the following are not true, do not calculate probability using (1).
  - Finite Ω: The number of possible outcomes is finite.
  - Equipossibility: The outcomes have equal probability of occurrence.
- The bases for identifying equipossibility were often
  - physical symmetry (e.g. a well-balanced die, made of homogeneous material in a cubical shape) or
  - a balance of information or knowledge concerning the various possible outcomes.
- Equipossibility is meaningful only for finite sample space, and, in this case, the evaluation of probability is accomplished through the definition of classical probability.

- We will NOT use this definition beyond this section. We will soon introduce a formal definition in Section 5.

Example 3.2 (Slide)**.** In drawing a card from a deck, there are 52 equally likely outcomes, 13 of which are diamonds. This leads to a probability of 13/52 or 1/4.

3.3. Basic properties of classical probability: From Definition 3.1, we can easily verified the properties below.

- $P(A) \geq 0$
- $P(\Omega) = 1$
- $P(\emptyset) = 0$
- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ which comes directly from
$$|A \cup B| = |A| + |B| - |A \cap B|.$$
- $A \perp B \Rightarrow P(A \cup B) = P(A) + P(B)$
- Suppose $\Omega = \{\omega_1, \ldots, \omega_n\}$ and $P(\{\omega_i\}) = \frac{1}{n}$. Then $P(A) = \sum\limits_{\omega \in A} P(\{\omega\})$.
    - The probability of an event is equal to the sum of the probabilities of its component outcomes because outcomes are mutually exclusive

3.4. In classical probability theory defined by Definition 3.1,

$$A \perp B \text{ is equivalent to } P(A \cap B) = 0.$$

However, In general probability theory, the above statement is NOT true.

Example 3.5 (Slides)**.** When rolling two dice, there are 36 (equiprobable) possibilities.

$$P[\text{sum of the two dice} = 5] = 4/36.$$

Example 3.6. *Chevalier de Mere's Scandal of Arithmetic*:

> Which is more likely, obtaining at least one six in 4 tosses of a fair die (event A), or obtaining at least one double six in 24 tosses of a pair of dice (event B)?

We have

$$P(A) = \frac{6^4 - 5^4}{6^4} = 1 - \left(\frac{5}{6}\right)^4 \approx .518$$

and

$$P(B) = \frac{36^{24} - 35^{24}}{36^{24}} = 1 - \left(\frac{35}{36}\right)^{24} \approx .491.$$

Therefore, the first case is more probable.

Remark 1: Probability theory was originally inspired by gambling problems. In 1654, Chevalier de Mere invented a gambling system which bet even money[3] on event B above. However, when he began losing money, he asked his mathematician friend Pascal to analyze his gambling system. Pascal discovered that the Chevalier's system would lose about 51 percent of the time. Pascal became so interested in probability and together with another famous mathematician, Pierre de Fermat, they laid the foundation of probability theory. [U-X-L Encyclopedia of Science]

Remark 2: de Mere originally claimed to have discovered a *contradiction in arithmetic*. De Mere correctly knew that it was advantageous to wager on occurrence of event A, but his experience as gambler taught him that it was not advantageous to wager on occurrence of event B. He calculated $P(A) = 1/6 + 1/6 + 1/6 + 1/6 = 4/6$ and similarly $P(B) = 24 \times 1/36 = 24/36$ which is the same as $P(A)$. He mistakenly claimed that this evidenced a contradiction to the arithmetic law of proportions, which says that $\frac{4}{6}$ should be the same as $\frac{24}{36}$. Of course we know that he could not simply add up the probabilities from each tosses. (By De Meres logic, the probability of at least one head in two tosses of a fair coin would be $2 \times 0.5 = 1$, which we know cannot be true). [20, p 3]

[3] Even money describes a wagering proposition in which if the bettor loses a bet, he or she stands to lose the same amount of money that the winner of the bet would win.

Definition 3.7. In the world of gambling, probabilities are often expressed by ***odds***. To say that the odds are n:1 *against* the event A means that it is n times as likely that A does not occur than that it occurs. In other words, $P(A^c) = nP(A)$ which implies $P(A) = \frac{1}{n+1}$ and $P(A^c) = \frac{n}{n+1}$.

"Odds" here has nothing to do with even and odd numbers. The odds also mean what you will win, in addition to getting your stake back, should your guess prove to be right. If I bet \$1 on a horse at odds of 7:1, I get back \$7 in winnings plus my \$1 stake. The bookmaker will break even in the long run if the probability of that horse winning is 1/8 (not 1/7). Odds are "even" when they are 1:1 - win \$1 and get back your original \$1. The corresponding probability is 1/2.

3.8. It is important to remember that classical probability relies on the assumption that the outcomes are ***equally likely***.

Example 3.9. ***Mistake*** made by the famous French mathematician Jean Le Rond d'Alembert (18th century) who is an author of several works on probability:

"The number of heads that turns up in those two tosses can be 0, 1, or 2. Since there are three outcomes, the chances of each must be 1 in 3."

4 Enumeration / Combinatorics / Counting

There are many probability problems, especially those concerned with gambling, that can ultimately be reduced to questions about cardinalities of various sets. ***Combinatorics*** is the *study of systematic counting methods*, which we will be using to find the cardinalities of various sets that arise in probability.

4.1. ***Addition Principle (Rule of sum)***:

- When there are m cases such that the ith case has n_i options, for $i = 1, \dots, m$, and no two of the cases have any options in common, the total number of options is $n_1 + n_2 + \cdots + n_m$.
- In set-theoretic terms, suppose that a finite set S can be partitioned into (pairwise disjoint parts) $S_1, S_2, \dots, S_m$. Then,

$$|S| = |S_1| + |S_2| + \cdots + |S_m|.$$

- The art of applying the addition principle is to partition the set S to be counted into "manageable parts"; that is, parts which we can readily count. But this statement needs to be qualified. If we partition S into too many parts, then we may have defeated ourselves. For instance, if we partition 8 into parts each containing only one element, then applying the addition principle is the same as counting the number of parts, and this is basically the same as listing all the objects of S. Thus, a more appropriate description is that the art of applying the addition principle is to partition the set S into not too many manageable parts.[1, p 28]

Example 4.2. [1, p 28] Suppose we wish to find the number of different courses offered by SIIT. We partition the courses according to the department in which they are listed. Provided there is no cross-listing (cross-listing occurs when the same course is listed by more than one department), the number of courses offered by SIIT equals the sum of the number of courses offered by each department.

Example 4.3. [1, p 28] A student wishes to take either a mathematics course or a biology course, but not both. If there are four mathematics courses and three biology courses for which the student has the necessary prerequisites, then the student can choose a course to take in $4 + 3 = 7$ ways.

4.4. ***Multiplication Principle (Rule of product)***:

- When a procedure can be broken down into m steps, such that there are n_1 options for step 1, and such that after the completion of step $i - 1$ $(i = 2, \ldots, m)$ there are n_i options for step i, the number of ways of performing the procedure is $n_1 n_2 \cdots n_m$.
- In set-theoretic terms, if sets $S_1, \ldots, S_m$ are finite, then $|S_1 \times S_2 \times \cdots \times S_m| = |S_1| \cdot |S_2| \cdot \cdots \cdot |S_m|$.
- For k finite sets $A_1, ..., A_k$, there are $|A_1| \cdots |A_k|$ k-tuples of the form $(a_1, \ldots, a_k)$ where each $a_i \in A_i$.

Example 4.5. Let A, B, and C be finite sets. How many triples are there of the form (a,b,c), where $a \in A$, $b \in B$, $c \in C$?

Example 4.6. Suppose that a deli offers three kinds of bread, three kinds of cheese, four kinds of meat, and two kinds of mustard. How many different meat and cheese sandwiches can you make?

First choose the bread. For each choice of bread, you then have three choices of cheese, which gives a total of $3 \times 3 = 9$ bread/cheese combinations (rye/swiss, rye/provolone, rye/cheddar, wheat/swiss, wheat/provolone ... you get the idea). Then choose among the four kinds of meat, and finally between the two types of mustard or no mustard at all. You get a total of $3 \times 3 \times 4 \times 3 = 108$ different sandwiches.

Suppose that you also have the choice of adding lettuce, tomato, or onion in any combination you want. This choice gives another 2 x 2 x 2 = 8 combinations (you have the choice "yes" or "no" three times) to combine with the previous 108, so the total is now $108 \times 8 = 864$.

That was the multiplication principle. In each step you have several choices, and to get the total number of combinations, multiply. It is fascinating how quickly the number of combinations grow. Just add one more type of bread, cheese, and meat, respectively, and the number of sandwiches becomes 1,920. It would take years to try them all for lunch. [15, p 33]

Example 4.7 (Slides)**.** In 1961, Raymond Queneau, a French poet and novelist, wrote a book called *One Hundred Thousand Billion Poems.* The book has ten pages, and each page contains a sonnet, which has 14 lines. There are cuts between the lines so that each line can be turned separately, and because all lines have the same rhyme scheme and rhyme sounds, any such combination gives a readable sonnet. The number of sonnets that can be obtained in this way is thus 10^{14} which is indeed a hundred thousand billion. Somebody has calculated that it would take about 200 million years of nonstop reading to get through them all. [15, p 34]

Example 4.8. [1, p 29–30] Determine the number of positive integers that are factors of the number

$$3^4 \times 5^2 \times 11^7 \times 13^8.$$

The numbers 3,5,11, and 13 are prime numbers. By the fundamental theorem of arithmetic, each factor is of the form

$$3^i \times 5^j \times 11^k \times 13^\ell,$$

where $0 \leq i \leq 4$, $0 \leq j \leq 2$, $0 \leq k \leq 7$, and $0 \leq \ell \leq 8$. There are five choices for i, three for j, eight for k, and nine for ℓ. By the multiplication principle, the number of factors is

$$5 \times 3 \times 8 \times 9 = 1080.$$

4.9. ***Subtraction Principle***: Let A be a set and let S be a larger set containing A. Then

$$|A| = |S| - |S \setminus A|$$

- When S is the same as Ω, we have $|A| = |S| - |A^c|$
- Using the subtraction principle makes sense only if it is easier to count the number of objects in S and in $S \setminus A$ than to count the number of objects in A.

4.10. ***Division Principle (Rule of quotient)***: When a finite set S is partitioned into equal-sized parts of m elements each, there are $\frac{|S|}{m}$ parts.

4.1 Four kinds of counting problems

4.11. Choosing objects from a collection is also called **sampling**, and the chosen objects are known as a **sample**. The four kinds of counting problems are [7, p 34]:

(a) ordered sampling of r out of n items with replacement: n^r;

(b) ordered sampling of $r \leq n$ out of n items without replacement: $(n)_r$;

(c) unordered sampling of $r \leq n$ out of n items without replacement: $\binom{n}{r}$;

(d) unordered sampling of r out of n items with replacement: $\binom{n+r-1}{r}$.

- See 4.20 for "bars and stars" argument.

4.12. Given a set of n distinct items/objects, select a distinct **ordered**[4] sequence (word) of length r drawn from this set.

(a) ***Ordered Sampling with replacement***: $\mu_{n,r} = n^r$

- Meaning

[4] Different sequences are distinguished by the order in which we choose objects.

- Ordered sampling of r out of n items with replacement.
    - An object can be chosen repeatedly.
- $\mu_{n,1} = n$
- $\mu_{1,r} = 1$
- Examples:
  - From a deck of n cards, we draw r cards with replacement; i.e., we draw each card, make a note of it, put the card back in the deck and re-shuffle the deck before choosing the next card. How many different sequences of r cards can be drawn in this way?

  - There are 2^r binary strings/sequences of length r.

(b) ***Ordered Sampling without replacement***:

$$\begin{aligned}(n)_r &= \prod_{i=0}^{r-1}(n-i) = \frac{n!}{(n-r)!}\\ &= \underbrace{n\cdot(n-1)\cdots(n-(r-1))}_{r \text{ terms}}; \quad r \le n\end{aligned}$$

- Meaning
  - Ordered sampling of $r \le n$ out of n items without replacement.
    - Once we choose an object, we remove that object from the collection and we cannot choose it again.
  - "the number of possible r-**permutations** of n distinguishable objects"
  - the number of sequences[5] of size r drawn from an alphabet of size n without replacement.

[5] Elements in a sequence are ordered.

- Example: $(3)_2 = 3 \times 2 = 6 =$ the number of sequence of size 2 drawn from an alphabet of size $= 3$ without replacement.

  Suppose the alphabet set is {A, B, C}. We can list all sequences of size 2 drawn from {A, B, C} without replacement:

  A B
  A C
  B A
  B C
  C A
  C B

- Example: From a deck of 52 cards, we draw a hand of 5 cards without replacement (drawn cards are not placed back in the deck). How many hands can be drawn in this way?

- For integers r, n such that $r > n$, we have $(n)_r = 0$.
- Extended definition: The definition in product form

$$(n)_r = \prod_{i=0}^{r-1}(n-i) = \underbrace{n \cdot (n-1) \cdots (n-(r-1))}_{\text{r terms}}$$

  can be extended to *any real number* n and a non-negative integer r. We define $(n)_0 = 1$. (This makes sense because we usually take the empty product to be 1.)
- $(n)_1 = n$
- $(n)_r = (n-(r-1))(n)_{r-1}$. For example, $(7)_5 = (7-4)(7)_4$.

- $(1)_r = \begin{cases} 1, & \text{if r} = 1 \\ 0, & \text{if r} > 1 \end{cases}$

Example 4.13. (Slides) Finger-Smudge on Touch-Screen Devices

Example 4.14. (Slides) *Probability of coincidence birthday*: Probability that there is at least two people who have the same birthday[6] in a group of r persons:

Example 4.15. It is surprising to see how quickly the probability in Example 4.14 approaches 1 as r grows larger.

Birthday Paradox: In a group of 23 randomly selected people, the probability that at least two will share a birthday (assuming birthdays are equally likely to occur on any given day of the year[7]) is about 0.5.

- At first glance it is surprising that the probability of 2 people having the same birthday is so large[8], since there are only 23 people compared with 365 days on the calendar. Some of the surprise disappears if you realize that there are $\binom{23}{2} = 253$ pairs of people who are going to compare their birthdays. [2, p. 9]

[6] We ignore February 29 which only comes in leap years.

[7] In reality, birthdays are not uniformly distributed. In which case, the probability of a match only becomes larger for any deviation from the uniform distribution. This result can be mathematically proved. Intuitively, you might better understand the result by thinking of a group of people coming from a planet on which people are always born on the same day.

[8] In other words, it was surprising that the size needed to have 2 people with the same birthday was so small.

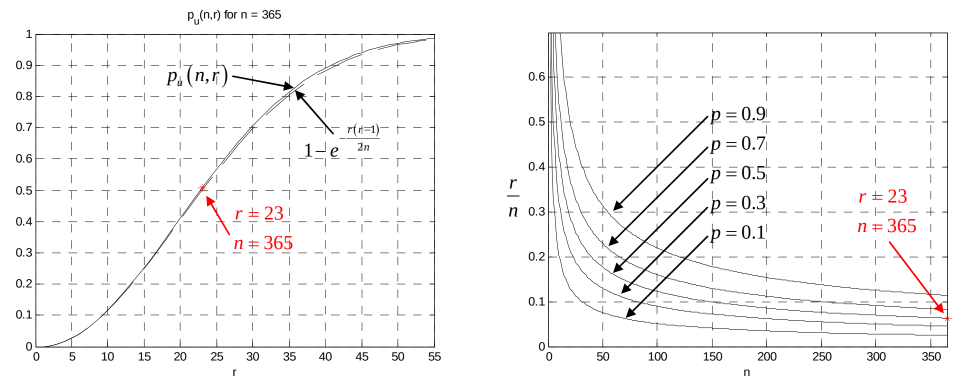

Figure 1: $p_u(n, r)$: The probability of the event that at least one element appears twice in random sample of size r with replacement is taken from a population of n elements.

Example 4.16. Another variant of the birthday coincidence paradox: The group size must be at least 253 people if you want a probability > 0.5 that someone will have the same birthday as you. [2, Ex. 1.13] (The probability is given by $1 - \left(\frac{364}{365}\right)^r$.)

- A naive (but incorrect) guess is that $\lceil 365/2 \rceil = 183$ people will be enough. The "problem" is that many people in the group will have the same birthday, so the number of different birthdays is smaller than the size of the group.

- On late-night television's The Tonight Show with Johnny Carson, Carson was discussing the birthday problem in one of his famous monologues. At a certain point, he remarked to his audience of approximately 100 people: "Great! There must be someone here who was born on my birthday!" He was off by a long shot. Carson had confused two distinctly different probability problems: (1) the probability of one person out of a group of 100 people having the same birth date as Carson himself, and (2) the probability of any two or more people out of a group of 101 people having birthdays on the same day. [20, p 76]

4.17. ***Factorial and Permutation***: The number of arrangements (permutations) of $n \geq 0$ distinct items is $(n)_n = n!$.

- For any integer n greater than 1, the symbol $n!$, pronounced "n factorial," is defined as the product of all positive integers less than or equal to n.
- $0! = 1! = 1$
- $n! = n(n-1)!$
- Computation:
  (a) `MATLAB`: Use `factorial(n)`. Since double precision numbers only have about 15 digits, the answer is only accurate for $n \leq 21$. For larger n, the answer will have the right magnitude, and is accurate for the first 15 digits.
  (b) Google's web search box built-in calculator: `n!`
- Meaning: The number of ways that n distinct objects can be ordered.
  - A special case of ordered sampling without replacement where $r = n$.
- In `MATLAB`, use `perms(v)`, where v is a row vector of length n, to creates a matrix whose rows consist of all possible permutations of the n elements of v. (So the matrix will contain $n!$ rows and n columns.)
- Example: In `MATLAB`, `perms([3 4 7])` gives

  7 4 3
  7 3 4
  4 7 3
  4 3 7
  3 4 7
  3 7 4

  Similarly, `perms('abcd')` gives

dcba dcab dbca dbac dabc dacb
cdba cdab cbda cbad cabd cadb
bcda bcad bdca bdac badc bacd
acbd acdb abcd abdc adbc adcb

4.18. ***Binomial coefficient***:

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

(a) Read "n choose r".

(b) Meaning:

(i) **Unordered sampling** of $r \leq n$ out of n items **without replacement**

(ii) The number of subsets of size r that can be formed from a set of n elements (without regard to the order of selection).

(iii) The number of combinations of n objects selected r at a time.

(iv) the number of k-**combinations** of n objects.

(v) The number of (unordered) sets of size r drawn from an alphabet of size n without replacement.

(c) Computation:

(i) `MATLAB`:

- `nchoosek(n,r)`, where n and r are nonnegative integers, returns $\binom{n}{r}$.

- `nchoosek(v,r)`, where v is a row vector of length n, creates a matrix whose rows consist of all possible combinations of the n elements of v taken r at a time. The matrix will contains $\binom{n}{r}$ rows and r columns.
  - Example: `nchoosek('abcd',2)` gives

    ab
    ac
    ad
    bc
    bd
    cd

(ii) Use `combin(n,r)` in `Mathcad`. However, to do symbolic manipulation, use the factorial definition directly.

(iii) In `Maple`, use $\binom{n}{r}$ directly.

(iv) Google's web search box built-in calculator: `n choose k`

(d) Reflection property: $\binom{n}{r} = \binom{n}{n-r}$.

(e) $\binom{n}{n} = \binom{n}{0} = 1$.

(f) $\binom{n}{1} = \binom{n}{n-1} = n$.

(g) $\binom{n}{r} = 0$ if $n < r$ or r is a negative integer.

(h) $\max\limits_{r} \binom{n}{r} = \binom{n}{\lfloor \frac{n+1}{2} \rfloor}$.

Example 4.19. In bridge, 52 cards are dealt to four players; hence, each player has 13 cards. The order in which the cards are dealt is not important, just the final 13 cards each player ends up with. How many different bridge games can be dealt? (Answer: 53,644,737,765,488,792,839,237,440,000)

4.20. The **bars and stars argument**:

- Example: Find all nonnegative integers x_1, x_2, x_3 such that

$$x_1 + x_2 + x_3 = 3.$$

$0+0+3$	1 1 1
$0+1+2$	1 1 1
$0+2+1$	1 1 1
$0+3+0$	1 1 1
$1+0+2$	1 1 1
$1+1+1$	1 1 1
$1+2+0$	1 1 1
$2+0+1$	1 1 1
$2+1+0$	1 1 1
$2+0+0$	1 1 1

- There are $\binom{n+r-1}{r} = \binom{n+r-1}{n-1}$ distinct vector $x = x_1^n$ of nonnegative integers such that $x_1 + x_2 + \cdots + x_n = r$. We use $n-1$ bars to separate r 1's.

  (a) Suppose we further require that the x_i are strictly positive ($x_i \geq 1$), then there are $\binom{r-1}{n-1}$ solutions.

  (b) ***Extra Lower-bound Requirement***: Suppose we further require that $x_i \geq a_i$ where the a_i are some given nonnegative integers, then the number of solution is

  $$\binom{r - (a_1 + a_2 + \cdots + a_n) + n - 1}{n-1}.$$

  Note that here we work with equivalent problem: $y_1 + y_2 + \cdots + y_n = r - \sum_{i=1}^{n} a_i$ where $y_i \geq 0$.

- Consider the distribution of $r = 10$ indistinguishable balls into $n = 5$ distinguishable cells. Then, we only concern with the number of balls in each cell. Using $n - 1 = 4$ bars, we can divide $r = 10$ stars into $n = 5$ groups. For example, ****|***||**|* would mean (4,3,0,2,1). In general, there are $\binom{n+r-1}{r}$ ways of arranging the bars and stars.

4.21. ***Unordered sampling with replacement***: There are n items. We sample r out of these n items with replacement. Because the order in the sequences is not important in this kind of sampling, two samples are distinguished by the number of each item in the sequence. In particular, Suppose r letters are drawn with replacement from a set $\{a_1, a_2, \ldots, a_n\}$. Let x_i be the number of a_i in the drawn sequence. Because we sample r times, we know that, for every sample, $x_1 + x_2 + \cdots x_n = r$ where the x_i are non-negative integers. Hence, there are $\binom{n+r-1}{r}$ possible unordered samples with replacement.

4.2 Binomial Theorem and Multinomial Theorem

4.22. ***Binomial theorem***: Sometimes, the number $\binom{n}{r}$ is called a ***binomial coefficient*** because it appears as the coefficient of $x^r y^{n-r}$ in the expansion of the binomial $(x+y)^n$. More specifically, for any positive integer n, we have,

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r} \tag{2}$$

(Slide) To see how we get (2), let's consider a smaller case of $n = 3$. The expansion of $(x+y)^3$ can be found using combinatorial reasoning instead of multiplying the three terms out. When $(x+y)^3 = (x+y)(x+y)(x+y)$ is expanded, all products of a term in the first sum, a term in the second sum, and a term in the third sum are added. Terms of the form x^3, x^2y, xy^2, and y^3 arise. To obtain a term of the form x^3, an x must be chosen in each of the sums, and this can be done in only one way. Thus, the x^3 term in the product has a coefficient of 1. To obtain a term of the form x^2y, an x must be chosen in two of the three sums (and consequently a y in the other sum). Hence. the number of such terms is the number of 2-combinations of three objects, namely, $\binom{3}{2}$. Similarly, the number of terms of the form xy^2 is the number of ways to pick one of the three sums to obtain an x (and consequently take a y from each of the other two terms). This can be done in $\binom{3}{1}$ ways. Finally, the only way to obtain a y^3 term is to choose the y for each of the three sums in the product, and this can be done in exactly one way. Consequently. it follows that

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

Now, let's state a combinatorial proof of the binomial theorem (2). The terms in the product when it is expanded are of the form $x^r y^{n-r}$ for $r = 0, 1, 2, \ldots, n$. To count the number of terms of the form $x^r y^{n-r}$, note that to obtain such a term it is necessary to choose r xs from the n sums (so that the other $n-r$ terms in the product are ys). Therefore. the coefficient of $x^r y^{n-r}$ is $\binom{n}{r}$.

From (2), if we let $x = y = 1$, then we get another important identity:

$$\sum_{r=0}^{n} \binom{n}{r} = 2^n. \tag{3}$$

4.23. *Multinomial Counting*: The ***multinomial coefficient*** $\binom{n}{n_1\ n_2\ \cdots\ n_r}$ is defined as

$$\prod_{i=1}^{r} \binom{n - \sum\limits_{k=0}^{i-1} n_k}{n_i} = \binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \binom{n-n_1-n_2}{n_3} \cdots \binom{n_r}{n_r} = \frac{n!}{\prod\limits_{i=1}^{r} n_i!}.$$

It is the number of ways that we can arrange $n = \sum\limits_{i=1}^{r} n_i$ tokens when having r types of symbols and n_i indistinguishable copies/tokens of a type i symbol.

4.24. *Multinomial Theorem*:

$$(x_1 + \ldots + x_r)^n = \sum \frac{n!}{i_1! i_2! \cdots i_r!} x_1^{i_1} x_2^{i_2} \cdots x_r^{i_r},$$

where the sum ranges over all ordered r-tuples of integers $i_1, \ldots, i_r$ satisfying the following conditions:

$$i_1 \geq 0, \ldots, i_r \geq 0, \quad i_1 + i_2 + \cdots + i_r = n.$$

When $r = 2$ this reduces to the binomial theorem.

4.3 Application: Success runs

Example 4.25. We are all familiar with "success runs" in many different contexts. For example, we may be or follow a tennis player and count the number of consecutive times the player's first serve is good. Or we may consider a run of forehand winners. A basketball player may be on a "**hot streak**" and hit his or her shots perfectly for a number of plays in row.

In all the examples, whether you should or should not be amazed by the observation depends on a lot of other information. There may be perfectly reasonable explanations for any particular success run. But we should be curious as to whether randomness could also be a perfectly reasonable explanation. Could the hot streak of a player simply be a snapshot of a random process, one that we particularly like and therefore pay attention to?

In 1985, cognitive psychologists Amos Taversky and Thomas Gilovich examined[9] the shooting performance of the Philadelphia 76ers, Boston Celtics and Cornell University's men's basketball team. They sought to discover whether a player's previous shot had any predictive effect on his or her next shot. Despite basketball fans' and players' widespread belief in hot streaks, the researchers found no support for the concept. (No evidence of nonrandom behavior.) [11, p 178]

4.26. Academics call the mistaken impression that a random streak is due to extraordinary performance the **hot-hand fallacy**. Much of the work on the hot-hand fallacy has been done in the context of sports because in sports, performance is easy to define and measure. Also, the rules of the game are clear and definite, data are plentiful and public, and situations of interest are replicated repeatedly. Not to mention that the subject gives academics a way to attend games and pretend they are working. [11, p 178]

Example 4.27. Suppose that two people are separately asked to toss a fair coin 120 times and take note of the results. Heads is noted as a "one" and tails as a "zero". The following two lists of compiled zeros and ones result

[9] "The Hot Hand in Basketball: On the Misperception of Random Sequences"

1 1 0 0 1 0 0 1 0 1 1 0 0 1 0 0 0 1 1 0 1 0 1 0 0 1 1 0 1 0
0 1 0 1 0 1 1 0 1 1 0 0 1 1 0 1 1 1 0 1 0 0 1 0 0 1 1 0 1 0
0 1 1 0 1 0 0 1 1 0 1 0 1 1 0 0 1 1 1 0 0 1 0 1 0 1 0 0 0 1
0 1 0 1 0 1 0 1 0 1 1 0 0 1 0 0 1 0 1 1 0 0 1 0 0 1 1 0 1 1

and

1 1 1 0 0 0 1 1 1 0 1 0 1 1 1 1 1 1 0 1 0 0 0 1 1 0 0 1 1 0
1 0 1 0 0 0 1 1 0 1 0 0 1 1 1 0 1 0 0 0 0 1 0 1 1 1 0 1 1 0
0 1 1 1 0 1 1 0 0 1 1 1 1 1 1 0 1 1 0 1 0 1 1 1 0 0 0 0 0 0
0 0 1 1 0 1 1 1 0 1 1 1 1 0 1 1 1 1 0 1 0 1 1 0 1 1 0 1 0 1

One of the two individuals has cheated and has fabricated a list of numbers without having tossed the coin. Which list is more likely be the fabricated list? [20, Ex. 7.1 p 42–43]

The answer is later provided in Example 4.33.

Definition 4.28. A **run** is a sequence of more than one consecutive identical outcomes, also known as a **clump**.

Definition 4.29. Let R_n represent the length of the longest run of heads in n independent tosses of a fair coin. Let $\mathcal{A}_n(x)$ be the set of (head/tail) sequences of length n in which the longest run of heads does not exceed x. Let $a_n(x) = \|\mathcal{A}_n(x)\|$.

Example 4.30. If a fair coin is flipped, say, three times, we can easily list all possible sequences:

HHH, HHT, HTH, HTT, THH, THT, TTH, TTT

and accordingly derive:

x	$P[R_3 = x]$	$a_3(x)$
0	1/8	1
1	4/8	4
2	2/8	7
3	1/8	8

4.31. Consider $a_n(x)$. Note that if $n \le x$, then $a_n(x) = 2^n$ because any outcome is a favorable one. (It is impossible to get more than three heads in three coin tosses). For $n > x$, we can partition

$\mathcal{A}_n(x)$ by the position k of the first tail. Observe that k must be $\le x+1$ otherwise we will have more than x consecutive heads in the sequence which contradicts the definition of $\mathcal{A}_n(x)$. For each $k \in \{1, 2, \dots, x+1\}$, the favorable sequences are in the form

$$\underbrace{\text{HH}\dots\text{H}}_{k-1 \text{ heads}}\text{T}\ \underbrace{\text{XX}\dots\text{X}}_{n-k \text{ positions}}$$

where, to keep the sequences in $\mathcal{A}_n(x)$, the last $n-k$ positions[10] must be in $\mathcal{A}_{n-k}(x)$. Thus,

$$a_n(x) = \sum_{k=1}^{x+1} a_{n-k}(x) \text{ for } n > x.$$

In conclusion, we have

$$a_n(x) = \begin{cases} \sum_{j=0}^{x} a_{n-j-1}(x), & n > x, \\ 2^n & n \le x \end{cases}$$

[19]. The following MATLAB function calculates $a_n(x)$

```
function a = a_nx(n,x)
a = [2.^(1:x) zeros(1,n-x)];
a(x+1) = 1+sum(a(1:x));
for k = (x+2):n
    a(k) = sum(a((k-1-x):(k-1)));
end
a = a(n);
```

4.32. Similar technique can be used to contract $\mathcal{B}_n(x)$ defined as the set of sequences of length n in which the longest run of heads and the longest run of tails do not exceed x. To check whether a sequence is in $\mathcal{B}_n(x)$, first we convert it into sequence of S and D by checking each adjacent pair of coin tosses in the original sequence. S means the pair have same outcome and D means they are different. This process gives a sequence of length $n-1$. Observe that a string of $x-1$ consecutive S's is equivalent to a run of length

[10]Strictly speaking, we need to consider the case when $n = x+1$ separately. In such case, when $k = x+1$, we have $\mathcal{A}_0(x)$. This is because the sequence starts with x heads, then a tail, and no more space left. In which case, this part of the partition has only one element; so we should define $a_0(x) = 1$. Fortunately, for $x \ge 1$, this is automatically satisfied in $a_n(x) = 2^n$.

x. This put us back to the earlier problem of finding $a_n(x)$ where the roles of H and T are now played by S and D, respectively. (The length of the sequence changes from n to $n-1$ and the max run length is $x-1$ for S instead of x for H.) Hence, $b_n(x) = \|\mathcal{B}_n(x)\|$ can be found by

$$b_n(x) = 2a_{n-1}(x-1)$$

[19].

Example 4.33. Continue from Example 4.27. We can check that in 120 tosses of a fair coin, there is a very large probability that at some point during the tossing process, a sequence of five or more heads or five or more tails will naturally occur. The probability of this is

$$\frac{2^{120} - b_{120}(4)}{2^{120}} \approx 0.9865.$$

0.9865. In contrast to the second list, the first list shows no such sequence of five heads in a row or five tails in a row. In the first list, the longest sequence of either heads or tails consists of three in a row. In 120 tosses of a fair coin, the probability of the longest sequence consisting of three or less in a row is equal to

$$\frac{b_{120}(3)}{2^{120}} \approx 0.000053,$$

which is extremely small indeed. Thus, the first list is almost certainly a fake. Most people tend to avoid noting long sequences of consecutive heads or tails. Truly random sequences do not share this human tendency! [20, Ex. 7.1 p 42–43]