ENGG2430A Probability and Statistics for Engineers

Chapter 1: Sample space and probability

Instructor: Shengyu Zhang

About the course

- Website: http://www.cse.cuhk.edu.hk/~syzhang/course/Prob17/
- You can find the lecture slides, tutorial slides, info for time, venue, TA, textbook, grading method, etc.
- No tutorial in the first week.
- Announcements will be posted on web.
  - The important ones will be sent to your cuhk email as well.

Content

- Sets.
- Probabilistic models.
- Conditional probability.
- Total Probability Theorem and Bayes' Rule.
- Independence.
- Counting.

Sets

- Probability makes extensive use of set operations.
- A set is a collection of objects, which are the elements of the set.
- $x \in S$: S is a set and x is an element of S
- $x \notin S$: x is not an element of S.
- $\emptyset$: A set that has no elements; called empty set.

Sets

- Subset: $S \subseteq T$
- Equal sets: $S = T$
- Countable vs. uncountable
- Universal set Ω: The set which contains all objects that could conceivably be of interest in a particular context.
- Complement: $\bar{S} = S^c = \Omega - S$.

Sets

- Union of sets: $S \cup T$, $\bigcup_{i=1}^{\infty} S_i$, $\bigcup_{i \in I} S_i$.
- Intersection of sets: $S \cap T$, $\bigcap_{i=1}^{\infty} S_i$, $\bigcap_{i \in I} S_i$.
- Disjoint sets: empty pairwise intersection.
- Partition of set S: a collection of disjoint sets whose union is S.
- De Morgan's laws:

$$\overline{\bigcup_i S_i} = \bigcap_i \overline{S_i}, \qquad \overline{\bigcap_i S_i} = \bigcup_i \overline{S_i}$$

Content

- Sets.
- Probabilistic models.
- Conditional probability.
- Total Probability Theorem and Bayes' Rule.
- Independence.
- Counting.

Experiment and outcomes

- A probabilistic model is a mathematical description of an uncertain situation.
- Every probabilistic model involves an underlying process, called the experiment.
  - *Example*. Flip two coins.
- The experiment produces exactly one out of several possible outcomes.
  - *Example*. four outcomes: $\{HH, HT, TH, TT\}$

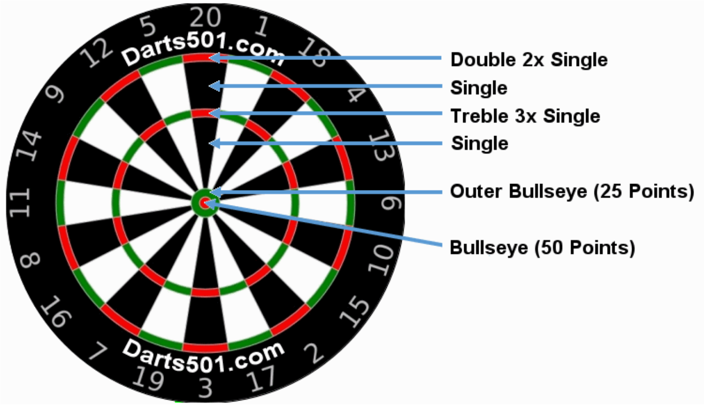
Sample space and events

- The set of all possible outcomes is the sample space, usually denoted by Ω.
  - *Example*. $\Omega = \{HH, HT, TH, TT\}$.

- Event: a subset of sample space.
  - $A \subseteq \Omega$ is a set of possible outcomes
  - *Example*. $A = \{HH, TT\}$, the event that the two coins give the same side.

Infinite sample space

- The sample space of an experiment may consist of a finite or an infinite number of possible outcomes.
  - Finite sample spaces are conceptually and mathematically simpler.
  - Sample spaces with an infinite number of elements are quite common.
    - As an example, consider throwing a dart on a board and viewing the point of impact as the outcome.
    - The region "Bullseye" is an event: it's a subset of the sample space.



Be careful with the sample space

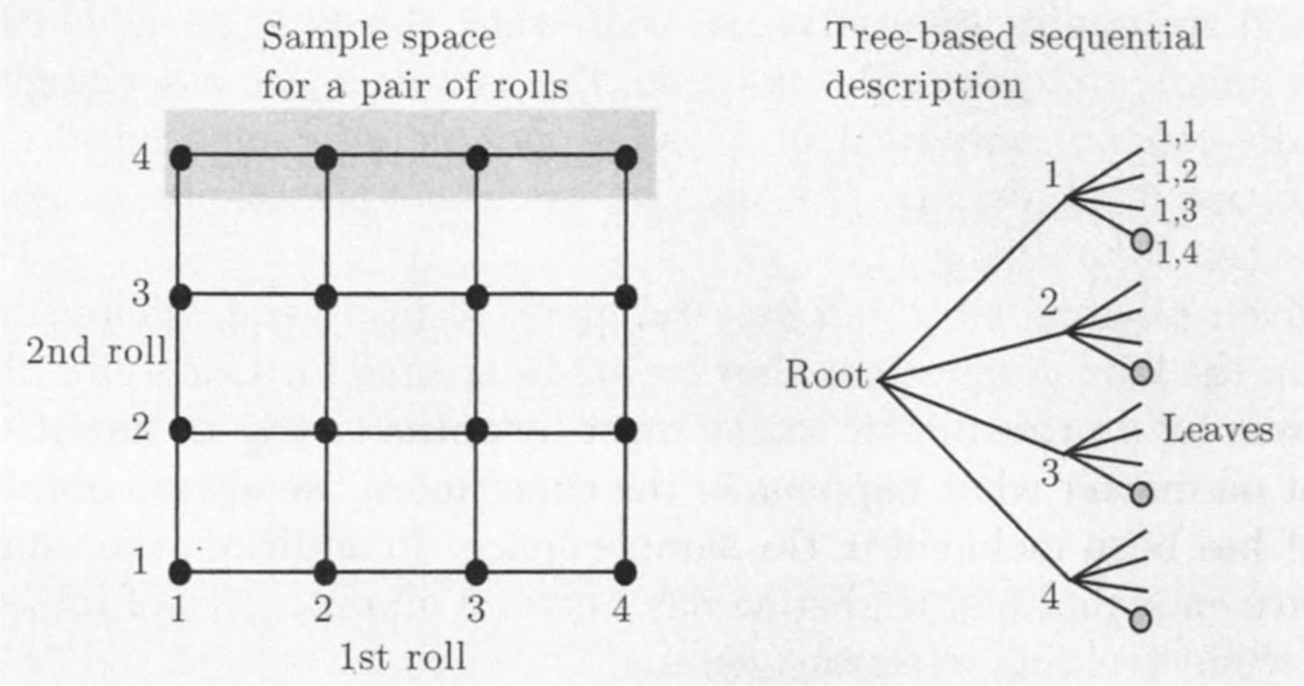
- One should choose an appropriate sample space.
- Different elements of the sample space should be distinct and mutually exclusive, so that when the experiment is carried out there is a unique outcome.
- The sample space must also be collectively exhaustive, in the sense that no matter what happens in the experiment, we always obtain an outcome that has been included in the sample space.

Sequential models

- Many experiments have an inherently sequential character.
- *Examples*:
  - tossing a coin three times,
  - observing the value of a stock on 5 successive days,
  - receiving eight successive digits at a communication receiver.
- It is then often useful to describe the experiment and the associated sample space by means of a tree-based sequential description.

Sequential models

- *Example*: row a 4-sided die twice.



Probabilistic laws

- After we settled on the sample space Ω associated with an experiment, we need to introduce a probabilistic law.
- The probability law assigns to a set A of possible outcomes a nonnegative number $P(A)$.
- The value $P(A)$ encodes our knowledge or belief about the collective "likelihood" of the elements of A.

Probabilistic laws

- Consider the *example* of tossing two coins.
- What's $P(HH)$? $P(HT)$? $P(TH)$? $P(TT)$?
  - Many possibilities. For example, uniform distribution says the following: $P(HH) = P(HT) = P(TH) = P(TT) = 1/4$.

- For $A = \{HH, TT\}$, what's $P(A)$?
- In uniform distribution, $P(A) = 1/2$.

Probability Axioms

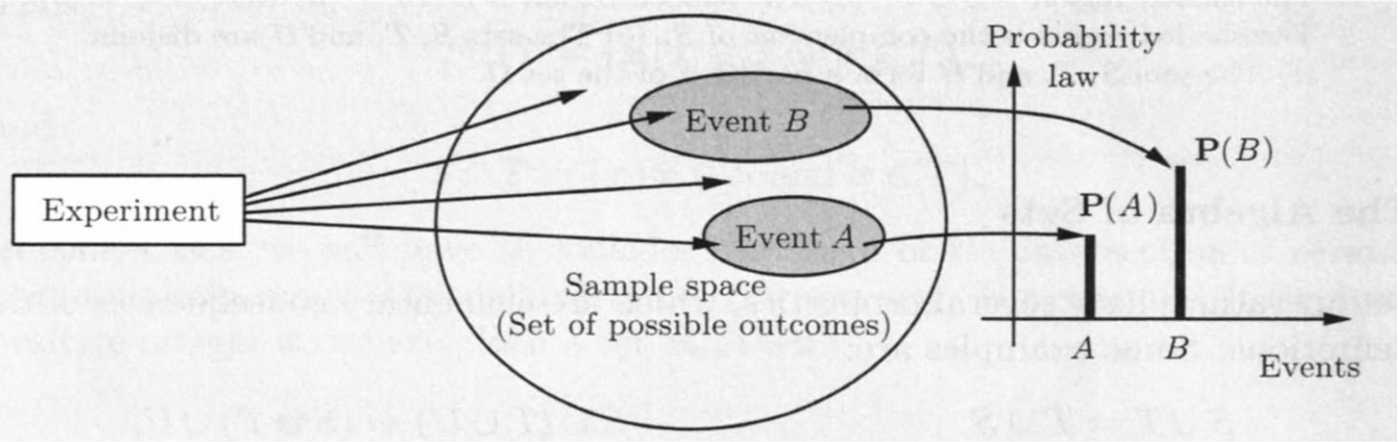
1. (*Non-negativity*) $P(A) \geq 0$, for every event A.

2. (*Additivity*) For any two disjoint events A and B, $P(A \cup B) = P(A) + P(B)$
   In general, if $A_1, A_2, \ldots$ are disjoint events, then
   $P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$

3. (*Normalization*) $P(\Omega) = 1$.

Probabilistic model: summary

- An experiment produces exactly one out of several possible outcomes.
- The sample space is the set of all possible outcomes.
- An event a subset of the sample space.
- The probability law assigns to any event A a number $P(A) \geq 0$.



Discrete Model

- In many cases, the sample space is discrete, and actually finite.
- Then the probability law is specified by the probabilities of the events that consist of a single element.
- It holds that for any event $A = \{a_1, \dots, a_n\}$,
$$P(A) = P(a_1) + \cdots + P(a_n).$$
- When the probability law is uniform, then
$$P(A) = |A|/|\Omega|.$$

Discrete model

- *Example*: toss a coin three times.
- The sample space is
- $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.
- Assume that each possible outcome has the same probability of 1/8.
- Consider event

  $A = \{exactly\ 2\ heads\ occur\} = \{HHT, HTH, THH\}$.
- $P(A) = P(\{HHT\}) + P(\{HTH\}) + P(\{THH\}) = 3/8$.

Continuous Model

- Sample space can also be infinite, and continuous.

- Caution: For continuous sample spaces, the probabilities of the single-element events may not be sufficient to characterize the probability law.

Continuous Model

- Consider $\Omega = [0,1]$.
  - Any number in the interval is a possible outcome.
- Assume uniform distribution: all outcomes happen equally likely.
- Then what's the probability of "½" as an outcome?
- What if you replace ½ with any of your favorite numbers?

Continuous Model

- Suppose the probability of a single element is $\varepsilon > 0$.
- No matter how small ε is, there is an integer $n > 0$, such that $1/n < \varepsilon$.
- Consider disjoint events $A_k = \{k/n\}$ for $k = 1,2, \ldots, n$.
- By additivity axiom

$$P(\Omega) \geq P(A_1) + P(A_2) + \cdots P(A_n) = n\varepsilon > 1,$$

  violating the rule that $P(\Omega) = 1$.
- Conclusion: $P(a) = 0$ for any outcome $a \in [0,1]$.
- So … what to do?

Continuous Model

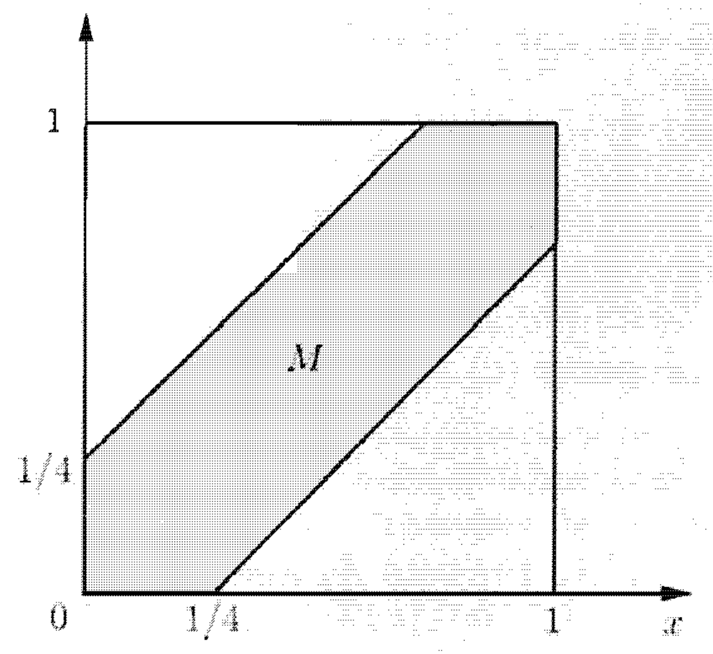
- A natural candidate: Define the probability on any subinterval $[a, b] \subseteq [0,1]$ to be
$$P([a, b]) = b - a$$
- Probability = "the length of the interval."
- And for disjoint union of intervals,
$$A = [a_1, b_1] \cup [a_2, b_2] \cup \cdots \cup [a_k, b_k] \cup \cdots,$$
define its probability by $P(A) = \sum_{i=1,2,\ldots,} (b_i - a_i)$
- Verify that all three axioms are satisfied.

Example: Meeting

- Romeo and Juliet have a date.
- Each will arrive at the meeting place with a delay between 0 and 1 hour, with all pairs of delays being equally likely.
- The first to arrive will wait for 15 minutes and will leave if the other has not yet arrived.
- *Question*: What is the probability that they will meet?

Example: Meeting

- Sample space: the unit square $[0,1] \times [0,1]$,
- Its elements are the possible pairs of delays.
- "equally likely" pairs of delays: let $P(A)$ for event $A \subseteq \Omega$ be equal to A's "area".
  - This satisfies the axioms.

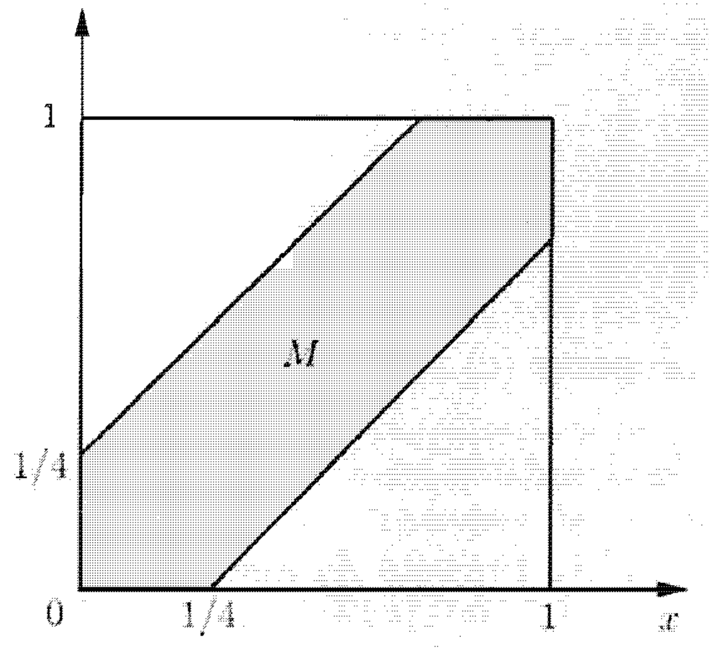


Example: Meeting

- The event that Romeo and Juliet will meet is the shaded region.
- Its probability is calculated to be 7/16.
    - $= 1 -$ the area of the two unshaded triangles

      $= 1 - 2 \cdot \left(\frac{3}{4}\right) \cdot \left(\frac{3}{4}\right) / 2$

      $= 7/16.$



$$M = \left\{ (x, y) \colon |x - y| \le \frac{1}{4}, 0 \le x \le 1, 0 \le y \le 1 \right\}$$

Properties of Probability Laws

- Consider a probability law, and let A, B, and C be events.

1. If $A \subseteq B$, then $P(A) \leq P(B)$.
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
3. $P(A \cup B) \leq P(A) + P(B)$.
4. $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$.
    - A^c is the complement of A.

Content

- Sets.
- Probabilistic models.
- Conditional probability.
- Total Probability Theorem and Bayes' Rule.
- Independence.
- Counting.

Partial information

- Conditional probability provides us with a way to reason about the outcome of an experiment, based on partial information.
- *Example*: In an experiment involving two successive rolls of a die, you are told that the sum of the two rolls is 9. How likely is it that the first roll was a 6?
- *Example*: How likely is it that a person has a certain disease given that a medical test was negative?
- *Example*: A spot shows up on a radar screen. How likely is it to correspond to an aircraft?

Conditional Probability

- In previous examples, we know that the outcome is within some given event B.
- We wish to quantify the likelihood that the outcome also belongs to some other event A.
- We seek to construct a new probability law that takes into account the available knowledge:
- a probability law that specifies the conditional probability of A given B.

Conditional Probability

- *Definition*. Conditional probability of A given B is
$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$
where we assume that $P(B) > 0$.
  - If $P(B) = 0$: then $P(A|B)$ is undefined.
- *Fact*. $P(A|B)$ form a legitimate probability law satisfying the three axioms.

Verification

1. Nonnegativity:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0.$$

2. Normalization:

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

3. Additivity: For two disjoint events A_1 and A_2, see the next slide.
   - The argument for a countable collection of disjoint sets is similar.

$$
\begin{aligned}
P(A_1 \cup A_2 | B) &= \frac{P\big((A_1 \cup A_2) \cap B\big)}{P(B)} \\
&= \frac{P\big((A_1 \cap B) \cup (A_2 \cap B)\big)}{P(B)} \\
&= \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)} \\
&= \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} \\
&= P(A_1|B) + P(A_2|B)
\end{aligned}
$$

Conditional probability: uniform case

- If the possible outcomes are finitely many and equally likely, then
$$P(A|B) = \frac{|A \cap B|}{|B|}.$$
- *Example 1*. Toss a fair coin three times.
- *Question*: What is the conditional probability $P(A|B)$ when A and B are:
    - $A = \{\text{more heads than tails come up}\}$
    - $B = \{1^{st} \text{ toss is a head}\}$

Conditional Probability: Example 1

- **Sample space:**

$$\Omega = \left\{ \begin{matrix} HHH, HHT, HTH, HTT, \\ THH, THT, TTH, TTT. \end{matrix} \right\}$$

- **Event $\boldsymbol{B}$** $= \{1^{st} \text{ toss is a head}\}$:

$$B = \{HHH, HHT, HTH, HTT\}$$

- **The probability of $\boldsymbol{B}$**:

$$P(B) = 4/8 = 1/2.$$

Conditional Probability: Example 1

- **Event $\boldsymbol{A \cap B}$:**

$$A \cap B = \{HHH, HHT, HTH\}$$

- **The probability of $\boldsymbol{A \cap B}$:**

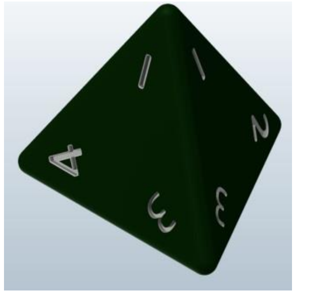
$$P(A \cap B) = 3/8.$$

- **The conditional probability:**

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3/8}{4/8} = \frac{3}{4}.$$

Conditional Probability: Example 2

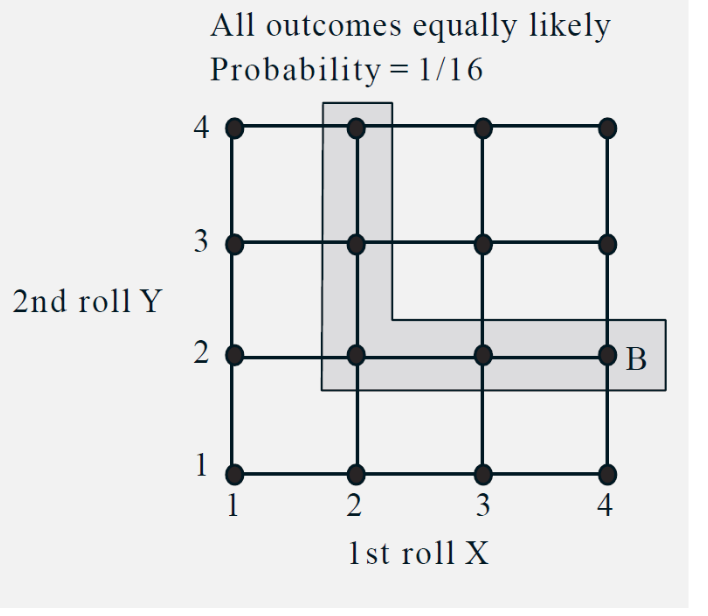
- Roll a fair 4-sided die twice
  - $X = outcome\ of\ 1^{st}\ roll$
  - $Y = outcome\ of\ 2^{nd}\ roll$



- The events A, B
  - $A = \{\max(X, Y) = m\} \quad m = 1,2,3,4$
  - $B = \{\min(X, Y) = 2\}$
- *Question*: What is the conditional probability $P(A|B)$?

Conditional Probability: Example 2

- Counting the number of elements of $A \cap B$ and B
  - $A = \{\max(X, Y) = m\}$
  - $B = \{\min(X, Y) = 2\}$



- $$P(A|B) = \begin{cases} 2/5, & \text{if } m = 3 \text{ or } m = 4 \\ 1/5, & \text{if } m = 2 \\ 0, & \text{if } m = 1 \end{cases}$$

Conditional Probability: Example 3

- Two teams N and C design a product within a month.
  - $P(C \text{ is successful}) = 2/3$
  - $P(N \text{ is successful}) = 1/2$
  - $P(\text{at least one team is successful}) = 3/4$
- *Question*: Assuming that exactly one successful design is produced, what is the probability that it was designed by team N?

Conditional Probability: Example 3

- 4 possible outcomes:

  SS: both succeed

  SF: C succeeds, N fails

  FF: both fail

  FS: C fails, N succeeds

- We know that

$$P(SS) + P(SF) = 2/3$$
$$P(SS) + P(FS) = 1/2$$
$$P(SS) + P(SF) + P(FS) = 3/4$$

And the normalization equation

$$P(SS) + P(SF) + P(FS) + P(FF) = 1$$

Conditional Probability: Example 3

- Solving the system of equations, we can obtain the probabilities of individual outcomes:

$$P(SS) = \frac{5}{12} \qquad P(SF) = \frac{1}{4}$$

$$P(FS) = \frac{1}{12} \qquad P(FF) = \frac{1}{4}$$

- The desired conditional probability is

$$P(FS|\{SF, FS\}) = \frac{\frac{1}{12}}{\frac{1}{4} + \frac{1}{12}} = \frac{1}{4}$$

Multiplication Rule

- *Fact*. Assuming that all of the conditioning events $A_1, A_1 \cap A_2, \ldots$ have positive probability, we have

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots \cdot P\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right)$$

Multiplication Rule: Example 1

- 3 cards drawn from 52-card deck without replacement.
  - drawn cards are not placed back in the deck.
- *Question*: What’s the probability that none of the three cards is a heart?
- One approach: count the number of card triplets that do not include a heart, and divide it with the number of all possible card triplets.
- Cumbersome.

Multiplication Rule: Example 1

- Another approach uses multiplication rule.
- $A_i = \{$the ith card is not a heart$\}$, $i = 1,2,3$.
- multiplication rule:
  $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$
- Since there are 39 cards that are not hearts,

$$P(A_1) = \frac{39}{52}$$

Multiplication Rule: Example 1

- Given that the first card is not a heart, we are left with 51 cards, 38 of which are not hearts:

$$P(A_2 | A_1) = \frac{38}{51}.$$

- Finally, given that the first two cards drawn are not hearts. there are 37 cards which are not hearts in the remaining 50 cards:

$$P(A_3 | A_1 \cap A_2) = \frac{37}{50}.$$

- Thus $P(A_1 \cap A_2 \cap A_3) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50} \approx 0.41.$

Multiplication Rule: Example 2

- 4 graduate and 12 undergraduate students are randomly divided into 4 groups of 4.
  - “randomly”: given assignment of some students to certain slots, any of the remaining students is equally likely to be assigned to any of the remaining slots.
- *Question*: What is the probability that each group includes a graduate student?

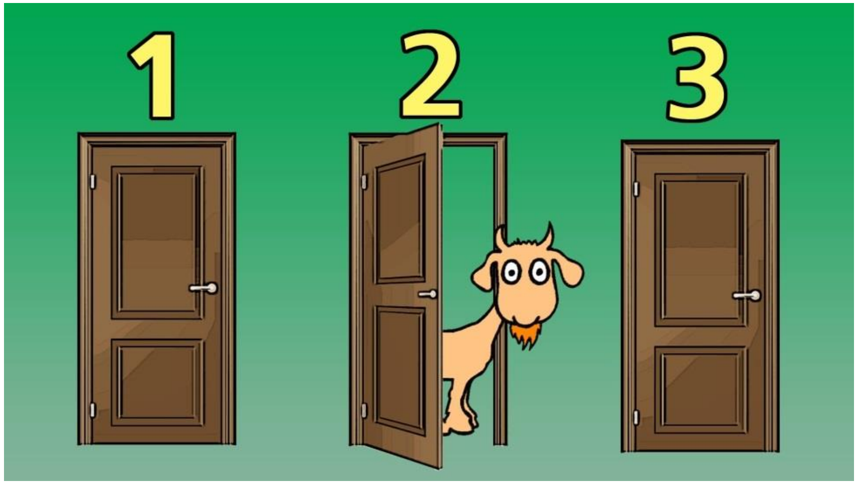
Multiplication Rule: Example 2

- Denote the four graduate students by $1, 2, 3, 4$
- Define events
  - $A_1 = \{$students 1 and 2 are in different groups$\}$,
  - $A_2 = \{$students 1, 2 and 3 are in different groups$\}$,
  - $A_3 = \{$students 1, 2, 3 and 4 are in different groups$\}$.
- We will use multiplication rule

$$P(A_3) = P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

- $P(A_1) = 12/15$, $P(A_2|A_1) = 8/14$,
  $P(A_3|A_1 \cap A_2) = 4/13$.
- So $P(A_3) = \frac{12}{15} \cdot \frac{8}{14} \cdot \frac{4}{13} \approx 0.14$.

The Monty Hall Problem



- A prize is randomly put behind one of the three closed doors.
- You point to one door.
- A friend opens one of the remaining two doors, after making sure that the prize is not behind it.
- *Question*: Should you stick to your initial choice, or switch to the other unopened door?

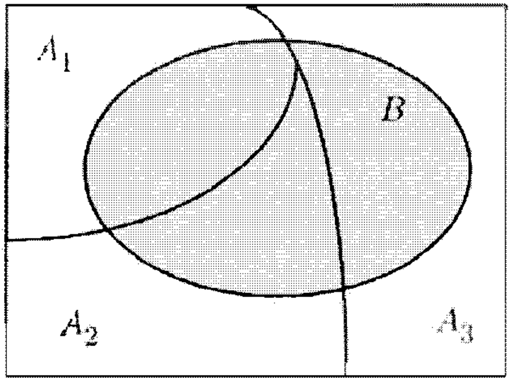
The Monty Hall Problem

- If sticking to the initial choice: the initial choice determines whether you win or not.
- Thus the winning probability is 1/3.
- If switching to the other unopened door:
  - Case 1: prize is behind the initial door, which happens with probability 1/3. You don’t win.
  - Case 2: prize is not behind the initial door, which happens with probability 2/3. You win for sure.
- So you should switch.

Content

- Sets.
- Probabilistic models.
- Conditional probability.
- Total Probability Theorem and Bayes' Rule.
- Independence.
- Counting.

Total Probability Theorem



- Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space. Assume $P(A_i) > 0$ for all i. Then, for any event B, we have

$$\begin{aligned} P(B) &= P(A_1 \cap B) + \cdots + P(A_n \cap B) \\ &= P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n) \end{aligned}$$

- Indeed, B is the the disjoint union of $(A_1 \cap B)$, …, $(A_n \cap B)$.
- The second equality is given by
$$P(A_i \cap B) = P(A_i)P(B|A_i).$$

Example: chess tournament

- Three types of players.
  - Type 1: 50%
  - Type 2: 25%
  - Type 3: 25%
- You winning probability with these players:
  - Against type 1: 0.3.
  - Against type 2: 0.4.
  - Against type 3: 0.5.
- Now you play a game with a randomly chosen player.
- *Question*: What’s your winning probability?

Example: chess tournament

- A_i: playing with an opponent of type i
  - $P(A_1) = 0.5,\ P(A_2) = 0.25, P(A_3) = 0.25.$
- B: winning
  - $P(B|A_1) = 0.3,\ \ P(B|A_2) = 0.4,\ \ P(B|A_3) = 0.5$

- The probability of B:

$$\begin{aligned} P(B) &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) \\ &= 0.50 \times 0.3 + 0.25 \times 0.4 + 0.25 \times 0.5 \\ &= 0.375 \end{aligned}$$

Example: Four-Sided Die

- Roll a fair 4-sided die.
- Rule: Roll once more if result is 1 or 2, otherwise stop

- *Question*: What is the probability that the sum total of your rolls is at least 4?

Example: Four-Sided Die

- A_i: the result of first roll is i
  $$P(A_i) = 1/4, \qquad \forall i = 1,2,3,4.$$
- B: the sum total is at least 4.
- $P(B) = \sum_{i=1}^{4} P(A_i)P(B|A_i)$. Let's calculate each $P(B|A_i)$.
- Given A_1: the sum total will be ≥ 4 if the second roll results in 3 or 4, which happens with probability $1/2$.
- Thus $P(B|A_1) = \frac{1}{2}$,
- Similarly $P(B|A_2) = \frac{3}{4}$.
  - Given A_2, the sum total will be ≥ 4 if the second roll results in 2, 3, or 4, which happens with probability $3/4$.

- Given A_3: you stop and the sum total remains below 4.
- Thus $P(B|A_3) = 0$,
- Given A_4: you stop but the sum total is already 4.
- Thus $P(B|A_4) = 1$.
- By the total probability theorem

$$P(B) = \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{3}{4} + \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 1 = \frac{9}{16}$$

Example: Up-to-date or Behind

- Alice is taking a probability class. At the end of each week,
  - she can be either up-to-date
  - or she may have fallen behind
- If she is up-to-date in week i, the probability that she will be up-to-date (or behind) in week $i + 1$ is 0.8 (or 0.2, respectively).
- If she is behind in a given week, the probability that she will be up-to-date (or behind) in the next week is 0.4 (or 0.6, respectively).
- Alice is up-to-date when she starts the class.

Example: Up-to-date or Behind

- *Question*: What is the probability that she is up-to-date after three weeks?
- U_i: Alice is up-to-date after i weeks
- B_i: Alice is behind after i weeks
- Previous slide:

$$P(U_{i+1}|U_i) = 0.8, \qquad P(U_{i+1}|B_i) = 0.4, \qquad P(U_0) = 1$$

- *Question* (rephrased): What is the probability of U_3?

Example: Up-to-date or Behind

- By total probability theorem

$$P(U_3) = P(U_2)P(U_3|U_2) + P(B_2)P(U_3|B_2)$$
$$= P(U_2) \cdot 0.8 + P(B_2) \cdot 0.4$$

- Similarly

$$P(U_2) = P(U_1) \cdot 0.8 \ + P(B_1) \cdot 0.4 = 0.72$$
$$P(B_2) = P(U_1) \cdot 0.2 \ + P(B_1) \cdot 0.6 = 0.28$$

- Since Alice starts her class up-to-date, we have

$$P(U_1) = 0.8, P(B_1) = 0.2.$$

- The probability of U_3

$$P(U_3) = \ 0.72 \cdot 0.8 \ + \ 0.28 \cdot 0.4 = 0.688$$

Bayes' Rule

- Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space, and assume that $P(A_i) > 0$, for all i.
- Then, for any event B with $P(B) > 0$, we have

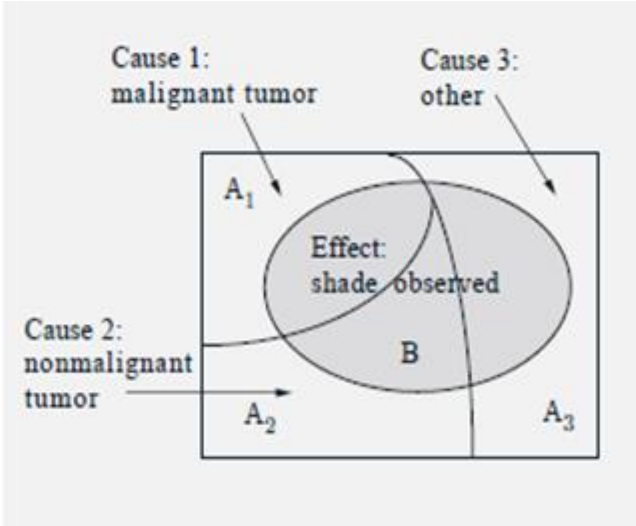
$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}$$

$$= \frac{P(A_i)P(B|A_i)}{P(B)}$$

$$= \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)}$$

Inference using Bayes' rule

- Bayes' rule is often used for inference.
- There are a number of causes that may result in a certain effect.
- We observe the effect and we wish to infer the cause.
- Causes: $A_1, \ldots, A_n$
- Effects: event B
  - $P(B|A_i)$: suppose known
  - $P(A_i|B)$: Posterior probability
  - $P(A_i)$: Prior probability



Example: Chess Revisited

- Three types of players.
  - Type 1: 50%
  - Type 2: 25%
  - Type 3: 25%
- You winning probability with these players:
  - Against type 1: 0.3.
  - Against type 2: 0.4.
  - Against type 3: 0.5.
- *Question*: Suppose that you win. What is the probability that you had an opponent of type 1?

Example: Chess Revisited

- A_i: getting an opponent of type i
  - $P(A_1) = 0.5,\ P(A_2) = 0.25, P(A_3) = 0.25.$
- B: the event of winning
  - $P(B|A_1) = 0.3,\ P(B|A_2) = 0.4,\ P(B|A_3) = 0.5$
- By Bayes' rule:

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1)+P(A_2)P(B|A_2)+P(A_3)P(B|A_3)}$$

$$= \frac{0.5 \cdot 0.3}{0.3 \cdot 0.5 + 0.25 \cdot 0.4 + 0.25 \cdot 0.5}$$

$$= 0.4$$

Example: Diagnosis

- A random person drawn from a certain population has probability 0.001 of having a certain disease.
- The test satisfies
  - $\Pr[\text{test positive} \mid \text{disease}] = 0.95$
  - $\Pr[\text{test negative} \mid \text{no disease}] = 0.95$
- *Question*: Given that the person just tested positive, what is the probability of having the disease?

Example: Diagnosis

- A: person has the disease
- B: test result is positive

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}$$

$$= \frac{0.001 \cdot 0.95}{0.001 \cdot 0.95 + 0.999 \cdot 0.05}$$

$$= 0.0187$$

- Much smaller than 95%!
- The Economist (February 20th, 1999): 80% of those questioned at a leading American hospital substantially missed the correct answer to a question of this type; most of them thought that the probability that the person has the disease is 0.95!

Content

- Sets.
- Probabilistic models.
- Conditional probability.
- Total Probability Theorem and Bayes' Rule.
- Independence.
- Counting.

Independence

- Consider two events A and B.
- A and B are independent: B provides no information of A.

$$P(A|B) = P(A)$$

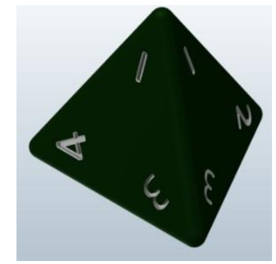
- Equivalently:

$$P(A \cap B) = P(A)P(B)$$

  - Why equivalent? $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Example: Dice rolling

- Consider the experiment of rolling a fair 4-sided dice twice.



- *Question*: Are the following events independent?
  - $A_i = \{1st \text{ roll results in } i\}$
  - $B_j = \{2nd \text{ roll results in } j\}$

Example: Dice rolling

- The probability of $A_i \cap B_j$:

$$P\left(A_i \cap B_j\right) = P(\text{the outcome of the two rolls is}(i,j)) = \frac{1}{16}$$

- The probability of A_i:

$$P(A_i) = \frac{\text{number of elements of } A_i}{\text{total number of possible outcomes}} = \frac{4}{16}$$

- The probability of B_j:

$$P\left(B_j\right) = \frac{\text{number of elements of } B_j}{\text{total number of possible outcomes}} = \frac{4}{16}$$

- Check the independence condition

$$P\left(A_i \cap B_j\right) = P(A_i)\, P(B_j)$$

- It holds, so the two events are independent.

Example: Dice rolling

- *Question*: Are the following events independent?
  - $A = \{1st \text{ roll is } 1\}$
  - $B = \{\text{sum of the two rolls is a } 5\}$
- The probability of $A \cap B$:

$$P(A \cap B) = P(1st \text{ roll is } 1, 2nd \text{ roll is } 4) = \frac{1}{16}$$

Example: Dice rolling

- The probability of A

$$P(A) = \frac{\text{number of elements of } A}{\text{total number of possible outcomes}} = \frac{4}{16}$$

- The probability of B

$$P(B) = \frac{\text{number of elements of } B}{\text{total number of possible outcomes}} = \frac{4}{16}$$

- Check the independence condition

$$P(A \cap B) = P(A)P(B)$$

- It holds, so the two events are independent.

Example: Dice rolling

- *Question*: Are the following events independent?
  - $A = \{\text{maximum of the two rolls is } 2\}$
  - $B = \{\text{minimum of the two rolls is } 2\}$
- The probability of $A \cap B$:

$$P(A \cap B) = P(\text{the result of the two rolls is } (2,2))$$
$$= \frac{1}{16}$$

Example: Dice rolling

- The probability of A

$$P(A) = \frac{\text{number of elements of } A}{\text{total number of possible outcomes}}$$
$$= \frac{|\{(1,2),(2,1),(2,2)\}|}{16} = \frac{3}{16}$$

- The probability of B

$$P(B) = \frac{\text{number of elements of } B}{\text{total number of possible outcomes}}$$
$$= \frac{|\{(2,2),(2,3),(2,4),(3,2),(4,2)\}|}{16} = \frac{5}{16}$$

- Check the independence condition and find

$$P(A \cap B) \neq P(A)P(B)$$

- Thus the two events are not independent. They are dependent.

Conditional independence

- Given an event C, the events A and B are *conditionally independent* if

$$P(A \cap B|C) = P(A|C) \cdot P(B|C)$$

- An equivalent formula is

$$P(A|B \cap C) = P(A|C)$$

- The equivalence is because

$$P(A \cap B \mid C) = \frac{P(A \cap B \cap C)}{P(C)} = \frac{P(C)P(B|C)P(A|B \cap C)}{P(C)}$$

$$= P(B|C)P(A|B \cap C)$$

Conditional independence: Example 1

- Consider two independent fair coin tosses
  - $A = \{1\text{st toss is a head}\}$
  - $B = \{2\text{nd toss is a head}\}$
  - $D = \{\text{the two tosses have different results}\}$
- Events A and B are independent, but

  $P(A|D) = \frac{1}{2},\ \ P(B|D) = \frac{1}{2},\ P(A \cap B|D) = 0.$
- Events A and B are not conditionally independent.

Conditional independence: Example 2

- Two biased coins, a blue one and a red one. Choose each with probability $1/2$.
- Blue coin: $P(H) = 0.99$
- Red coin: $P(H) = 0.01$
- Consider the events
  - $A = \{\text{1st toss results in head}\}$
  - $B = \{\text{2nd toss results in head}\}$
  - $D = \{\text{the blue coin is selected}\}$

Conditional independence: Example 2

- No matter which coin is chosen, the two tosses are independent.
- Namely, conditioned on D, A and B are independent.
- The probability of $A \cap B$ conditioned on D:

$$P(A \cap B|D) = P(A|D)P(B|D) = 0.99 \times 0.99$$

Conditional independence: Example 2

- The probability of A:

$$P(A) = P(D)P(A|D) + P(D^c)P(A|D^c) = 1/2$$

- Similarly, we have $P(B) = 1/2$.
- Check the independence condition

$$P(A \cap B) = P(D)P(A \cap B|D) + P(D^c)P(A \cap B|D^c)$$

$$= \frac{1}{2} \times 0.99 \times 0.99 + \frac{1}{2} \times 0.01 \times 0.01 \cong \frac{1}{2}$$

$$\neq P(A)P(B)$$

- Thus without the condition, A and B are dependent.

Independence of many events

- We say that the events $A_1, A_2, \ldots, A_n$ are independent if for every subset S of $\{1,2,\ldots,n\}$,

$$P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i)$$

- Note, pairwise independence does not imply independence.

Content

- Sets.
- Probabilistic models.
- Conditional probability.
- Total Probability Theorem and Bayes' Rule.
- Independence.
- Counting.

Counting

- The calculation of probabilities often involves counting the number of outcomes in various events.
  - Uniform distribution over finite sample space:
    $$P(A) = \frac{|A|}{|\Omega|}$$
  - An event A with a finite number of equally likely outcomes, each of which has probability p:
    $$P(A) = p \cdot |A|.$$

Combinatorics

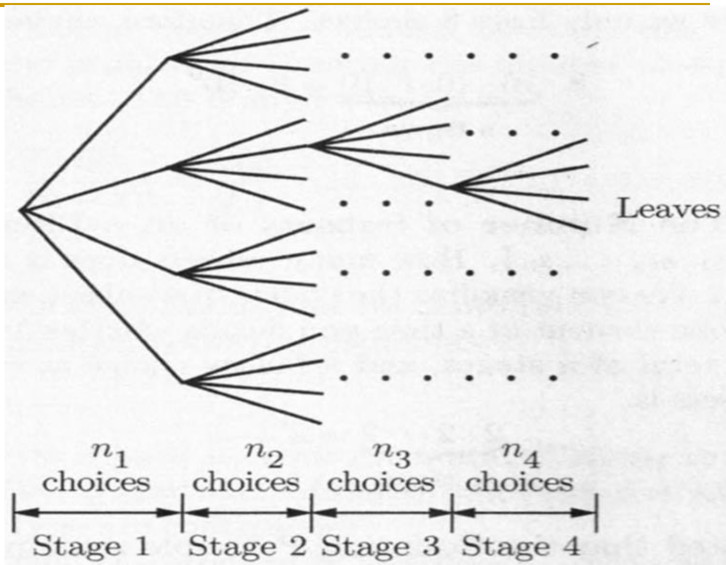
- The art of counting constitutes a large portion of the field of combinatorics.

- Next:
  - present the basic principle of counting
  - apply it to a number of situations that are often encountered in probabilistic models.

2 stages

- Consider an experiment that consists of two consecutive stages.
- The possible results at the first stage are $a_1, a_2, \dots, a_m$.
- The possible results at the second stage are $b_1, b_2, \dots, b_n$.
- Then the possible results of the two-stage experiment are all possible ordered pairs (a_i, b_j), $i = 1, \dots, m,\ j = 1, \dots, n$.
- The number of such ordered pairs: mn.

Multiple stages



- And this easily extends to multiple stages.
- Suppose r stages
  - There are n_1 possible results at the first stage.
  - For every possible result at the first stage, there are n_2 possible results at the second stage.
  - More generally, for any sequence of possible results at the first $i-1$ stages, there are n_i possible results at the ith stage.
- Then the total number of possible results of the r-stage process is $n_1 n_2 \ldots n_r$.

Example: Tel numbers

- A local telephone number is a 8-digit sequence, but the first digit has to be different from 0 or 1.
- *Question*: How many distinct telephone numbers are there?
- We have a total of 8 stages,
  - the first stage we only have 8 choices.
  - For the rest stages we have a 10 choices
- Therefore, the answer is

$$8 \times 10^7$$

Example: number of subsets

- Consider an n-element set $\{s_1, \ldots, s_n\}$
- *Question*: How many subsets does it have?
  - including itself and the empty set
- We can visualize the choice of a subset as a sequential process
  - examine one element at a time and decide whether to include it in the set or not.
- A total of n stages, and a binary choice at each stage.
- Therefore the number of subsets is 2^n.

k-permutations

- We start with n distinct objects, and let k be some positive integer, with $k \leq n$.

- We wish to count the number of different ways that we can pick k out of these n objects and arrange them in a sequence,
  - i.e., the number of distinct k-object sequences.

- We can choose any of the n objects to be the first one.
- Having chosen the first, there are only $n-1$ possible choices for the second.
- Given the choice of the first two, there only remain $n-2$ available objects for the third stage, etc.
- When we are ready to select the last (the kth) object, we have already chosen $k-1$ objects, which leaves us with $n-(k-1)$ choices for the last one.

- The number of possible sequences, called k-permutations, is

$$n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

- In the case of $k = n$, the number of possible sequences, called permutations, is

$$n(n-1)\ldots(n-k+1) = n!$$

  - Convention: $0! = 1$.

- *Question*: What's the number of words that consist of four distinct letters?
- This is the problem of counting the number of 4-permutations of the 26 letters in the alphabet.
- The number is

$$\frac{26!}{22!} = 26 \times 25 \times 24 \times 23 = 358{,}800$$

Combination

- There are n people and we are interested in forming a committee of k. How many different committees are possible?
- More abstractly, this is the same as the problem of counting the number of k-element subsets of a given n-element set.
- Forming a combination is different than forming a k-permutation, because in a combination there is no ordering of the selected elements.

- For example, whereas the 2-permutations of the letters A, B, C, and D are

  AB, BA, AC, CA, AD, DA, BC, CB, BD, DB, CD, DC,

- The combinations of two out of these four letters are

  AB, AC, AD, BC, BD, CD.

- In the preceding example, the combinations are obtained from the permutations by grouping together "duplicates“.
- For example, AB and BA are not viewed as distinct, and are both associated with the combination AB.
- In general, each combination is associated with $k!$ “duplicate” k-permutations, so the number $n!/(n - k)!$ of k-permutations = the number of combinations times $k!$.

- Hence, the number of possible combinations, is equal to $\frac{n!}{k!(n-k)!}$
- This is the same as the binomial coefficient $\binom{n}{k}$.

Example: an algebraic identity

- We have a group of n persons.
- Consider clubs that consist of a special person from the group (the club leader) and a number (possibly zero) of additional club members.
- Let us count the number of possible clubs of this type in two different ways, thereby obtaining an algebraic identity.

Method 1

- There are n choices for club leader.
- Once the leader is chosen, we are left with a set of $n-1$ available persons, and we are free to choose any of the 2^{n-1} subsets.
- Thus the number of possible clubs is $n2^{n-1}$.

Method 2

- For fixed k, we can form a k-person club by first selecting k out of the n available persons
  - There are $\binom{n}{k}$ choices.
- We can then select one of the members to be the leader (there are k choices).
- By adding over all possible club sizes k, we obtain the number of possible clubs as $\sum_{k=1}^{n} k\binom{n}{k}$.
- We thus showed the identity $\sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}$.

Partitions

- We are given an n-element set S and integers $n_1, \dots, n_r$.
  - $n_i \geq 0, \forall i \in \{1, \dots, r\}$
  - $n_1 + \cdots + n_r = n$.
- Task: Partition the set S into r disjoint subsets,
  - with the i-th subset containing exactly n_i elements.
- *Question*: How many ways can this be done?

- We form the subsets one at a time.
- We have $\binom{n}{n_1}$ ways of forming the first subset.
- Having formed the first subset, to form the second subset,
  - we are left with $n - n_1$ elements,
  - and need to choose n_2 of them.
- We have $\binom{n-n_1}{n_2}$ choices.
- Similar treatment for the rest…
- Counting Principle: total number of choices is

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-\cdots-n_{r-1}}{n_r}$$

Simplification

- $\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\dots\binom{n-n_1-\cdots-n_{r-1}}{n_r}$

$$= \frac{n!}{n_1!\,(n-n_1)!}\cdot\frac{(n-n_1)!}{n_2!\,(n-n_1-n_2)!}\cdot\dots$$

$$\cdot\frac{(n-n_1-\cdots-n_{r-1})!}{n_r!\,(n-n_1-n_2-\cdots-n_r)!}$$

$$= \frac{n!}{n_1!\,n_2!\dots n_r!}$$

- This is the same as the multinomial coefficient $\binom{n}{n_1,n_2,\dots,n_r}$.

Example: Anagrams

- *Question*: How many different words (letter sequences) can be obtained by rearranging the letters in the word TATTOO?
- There are 6 positions to be filled by the available letters.
- Each rearrangement corresponds to a partition of the set of the 6 positions into
  - a group of size 3: the positions that get the letter T
  - a group of size 1: the position that gets the letter A
  - a group of size 2: the positions that get the letter O

- Thus, the desired number is

$$\frac{6!}{1!\,2!\,3!} = 60.$$

Example: Students grouping (again)

- A class consisting of 4 graduate and 12 undergraduate students is randomly divided into four groups of 4.
  - “Randomly”: All partitions (into 4 groups of size 4) occur equally likely.
- *Question*: What is the probability that each group includes a graduate student?
- We’ve seen this before, but we’ll now obtain the answer using a different argument.

- Sample space Ω: All partitions of the 16 students into 4 groups of size 4.
- The size of the sample space:

$$|\Omega| = \binom{16}{4,4,4,4} = \frac{16!}{4!\,4!\,4!\,4!}$$

- Consider the event of each group containing a graduate student.
- Two steps: first allocate the graduate students, and then the undergraduate ones.

Allocation of grads

- There are
  - four choices for the group of the first graduate student,
  - three choices for the second,
  - two for the third,
  - one for the fourth.
- Thus, there is a total of 4! choices for this step.

Allocation of under

- Take the remaining 12 undergraduate students and distribute them to the four groups
  - 3 students in each.
- This can be done in

$$\binom{12}{3,3,3,3} = \frac{12!}{3!\,3!\,3!\,3!}$$

different ways.

- By the Counting Principle, the event of interest can occur in

$$\frac{4!\,12!}{3!\,3!\,3!\,3!}$$

different ways.

- The probability of this event is thus

$$\left(\frac{4!\,12!}{3!\,3!\,3!\,3!}\right) / \left(\frac{16!}{4!\,4!\,4!\,4!}\right) = \frac{12 * 8 * 4}{15 * 14 * 13}$$

same as previously calculated.