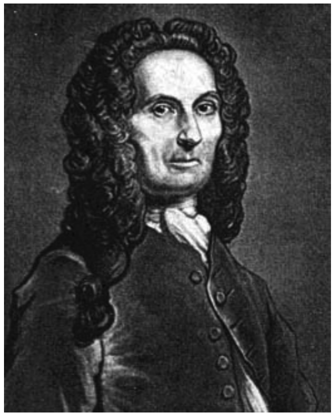
Chapter 2

Probability

"Chance, too, which seems to rush along with slack reins, is bridled and governed by law" **(Boethius, 480-524).**

Blaise Pascal (1623-1662)

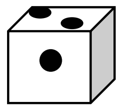
Pierre de Fermat (1601-1665)



Abraham de Moivre (1667-1754)

2.1 Sample space and events*

Experiment: any action or process whose outcome is uncertain.

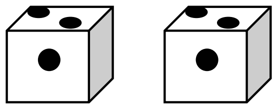
Example 2.1 *Roll a die.* 

Outcome: A particular realization of a random experiment. It is usually denoted by $\omega_1, \omega_2, \cdots$.

In the example the outcome could be landing with 1, 2, 3, 4, 5, or 6 on the top, e.g. $\omega_2 = 2$,

Sample space: The set (collection) of all possible outcomes, denoted by Ω. For intance $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Event: A set (collection) of outcomes or equivalently, a subset of the sample space. Usually denoted by $A, B, C, \cdots$. E.g. $A = \{2, 5\}$, $B = \{6\}, C = \emptyset$.

Question: What is the sample space of rolling two dice: 

Event (Set) Basic operations:

♠ **Union** $A \bigcup B = A$ or B — either A or B occurs (or both)

♠ **Intersection** $A \bigcap B = A$ and B — both A and B occur.

♠ **Complement** $A^c = \Omega - A$ — event A does not occur ($\overline{A}$)

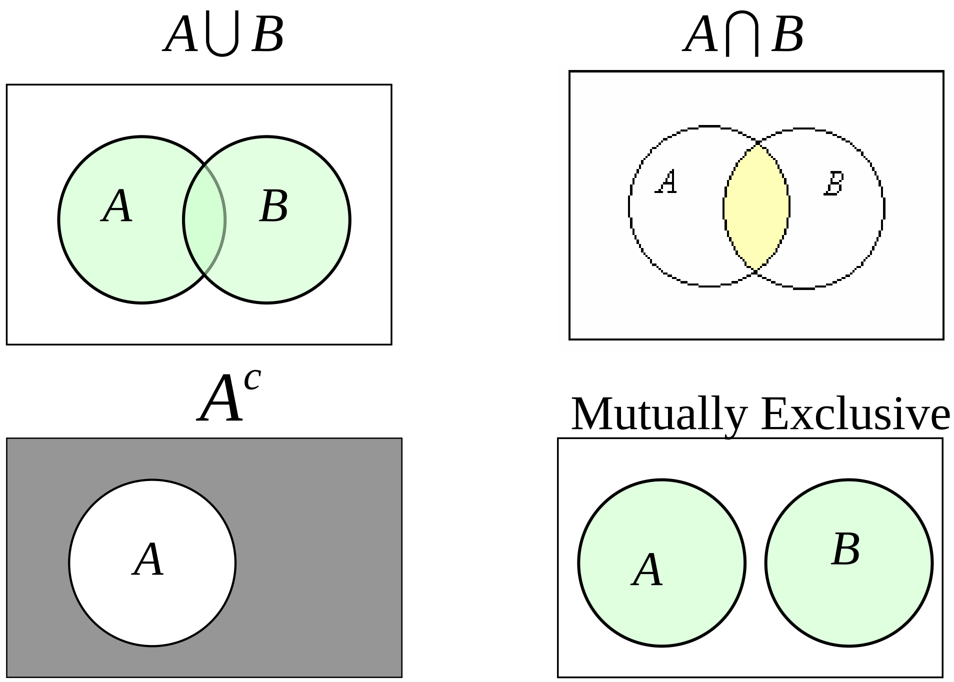


Figure 2.1: The Venn diagram shows the relationship between two sets

e.g. $A = \{1, 2\}, \quad B = \{2, 5\}$, then $A \bigcup B = \{1, 2, 5\}$, $A \bigcap B = \{2\}$, $A^c = \{3, 4, 5, 6\}$.

Mutually exclusive: A and B have no outcomes in common, written as $A \cap B = \emptyset$, or more generally $A_1, \cdots, A_n$ share **no common outcomes**.

Notation:

$$A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^n A_i$$

$$A \cap B = AB$$

$$A_1 \cap A_2 \cap \cdots \cap A_n = A_1 A_2 \cdots A_n = \bigcap_{i=1}^n A_i$$

de Morgan rule:

$$\left(\bigcup_{i=1}^n A_i\right)^c = \bigcap_{i=1}^n A_i^c \text{ and } \left(\bigcap_{i=1}^n A_i\right)^c = \bigcup_{i=1}^n A_i^c.$$

Example 4. Let Ω be the salaries earned by the graduates from a UK Business School in the past 5 years (surveyed in 2010). We may choose $\Omega = [0, \infty)$. Based on the dataset "Jobs.txt", we extract some *interesting* events/subsets. Recall the information of the dataset:

C1: ID number

C2: Job type, 1 - accounting, 2 - finance, 3 - management, 4 - marketing and sales, 5 -others

C3: Sex, 1 - male, 2 - female

C4: Job satisfaction, 1 - very satisfied, 2 - satisfied, 3 - not satisfied

C5: Salary (in thousand pounds)

C6: No. of jobs after graduation

```
IDNo.   JobType  Sex    Satisfaction  Salary     Search
1          1       1      3              51       1
2          4       1      3              38       2
3          5       1      3              51       4
4          1       2      2              52       5
... ...
```

Male and female salaries can be extracted as follows:

```
> jobs <- read.table("Jobs.txt", skip=10, header=T, row.names=1)
> mSalary <- jobs[,4][jobs[,2]==1]          #male salary
> fSalary <- jobs[jobs[,2]==2,4]            #female salary
```

We may extract the salaries from finance sector or accounting:

```
> finSalary <- jobs[,4][jobs[,1]==2]; summary(finSalary)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  37.00   46.00   53.00   52.08   58.00   65.00
> accSalary <- jobs[,4][jobs[,1]==1]; summary(accSalary)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  40.00   47.00   51.00   50.45   54.00   62.00
```

According to this dataset, finance pays slightly higher than accounting. We may also extract the salaries for males (females) in accounting:

```
> maccSalary <- jobs[,4][(jobs[,1]==1) & (jobs[,2]==1)]
      # ‘&’ stands for logic operation ‘and’
> summary(mfinSalary)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  44.00   48.00   51.00   51.31   55.00   62.00
```

```
> faccSalary<- jobs[,4][(jobs[,1]==1) & (jobs[,2]==2)]
> summary(ffinSalary)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  40.00   45.25   49.50   49.66   53.00   61.00
```

To extract the salaries for males in both finance and accounting:

```
>  mfinaccSalary <- jobs[,4][ (jobs[,2]==1) & ( (jobs[,1]==1) |
       (jobs[,1]==2) ) ]     # '|' stands for logic operation 'or'
```

To remove (unwanted) objects:

```
> rm(mSalary, fSalary, accSalary, finSalary, maccSalary,
          faccSalary, mfinaccSalary)
```

2.2 Probability

Probability shows the likelihood that an event can occur.

Two viewpoints of probability:

♣ **Frequentist**: prob = long-run relative freq (appealing for repeatable experiments).

♣ **Bayesian**: probability = measure of prior belief (appealing for non-repeatable experiments).

For example, when rolling a die, if $A = \{1, 2\}$, then $P(A) = 2/6 = 1/3$. If $B = \{2, 5\}$, then $P(A \cup B) = 3/6$ and $P(AB) = 1/6$.

Axioms of Probability (Komolgorov):

1. $P(A) \geq 0$.

2. $P(\Omega) = 1$.

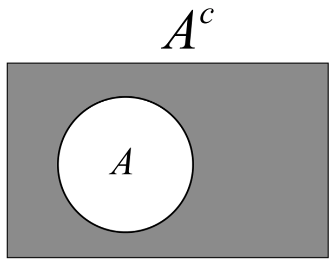
3. **Additivity**: If $A_1, A_2, \cdots$ are mutually exclusive, then

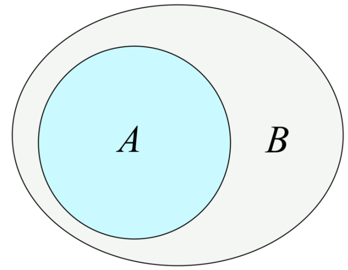
   $$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots.$$

All the other probability properties that you might know are derived from those 3 simple axioms.

Properties of Probability:

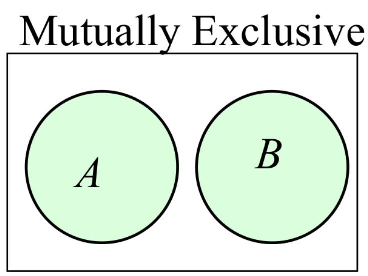
- (**Complement**) $P(A^c) = 1 - P(A)$.

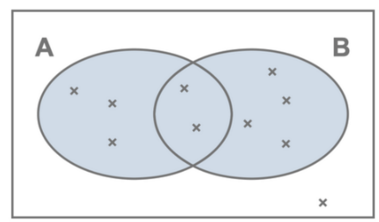




- (**Monotonicity**) if $A \subseteq B$ then $P(A) \leq P(B)$.

- If A and B are mutually exclusive, then $P(A \cap B) = 0$.





- (**Inclusion-exclusion**) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

- (**Union Bound**) $P(A \cup B) \leq P(A) + P(B)$.

Equally Likely Outcomes: $P(A) = \frac{\#\text{ elementary outcomes in } A}{\#\text{ elementary outcomes in } \Omega}$.

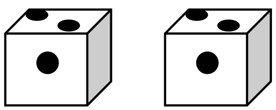
Example 2.2 *A machine tosses 5 coins at random. What is the probability of at least 4 heads?*

Possible outcomes $= \{HHHHH, THHHH, HTHHH, HHTHH, HHHTH, HHHHT, HHHTT, \cdots\}$, $2^5 = 32$ possible outcomes.

$$\begin{aligned} P(\text{at least 4 heads}) &= P(\text{exactly 4 heads}) + P(\text{exactly 5 heads}) \\ &= \frac{5}{32} + \frac{1}{32} = \frac{3}{16}. \end{aligned}$$

2.3 Counting techniques*

Rule of product: When an experiment consists of two parts, the first part of m distinct outcomes and the second part of n results, then we have total $m \times n$ outcomes. This **generalizes** to k-tuples.

Example 2.3 *Two dice are rolled.* 

- ♠ How many possible results?
- ♠ What is the chance that the sum of two dice is more than 7? Answer: $\frac{0+1+2+3+4+5}{36} = \frac{5}{12}$.

Example 2.4 *Rule of Permutations*

A class consists of 110 students. In how may ways can the first, second and third prize be awarded to the students?

First		Second		Third	
110	×	109	×	108	= 1,294,920

In this case, the ordering is important. For example, Jane, Richard and Caroline are different from Richard, Caroline and Jane for the first, second and third prize.

Rule of Permutations: Any **ordered** sequence of k objects taken from a set of n distinct objects is called a **permutation** (or an *arrangement*). The total number of permutations is

$$P_{k,n} = n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!},$$

where $m! = m(m-1)\cdots 1$ is the 'm factorial'.

Example (continued) In how many ways can three students be selected to form a committee? In this case, the ordering does not matter, namely, Jane, Richard and Caroline is the same committee as Richard, Caroline and Jane. There are 3! arrangements that correspond to the same committee. Hence, the total number is

$$\frac{110 \times 109 \times 108}{3 \times 2 \times 1} = \frac{110!}{3! \times 107!} = 215,820.$$

Rule of Combinations: The number of combinations obtained when selecting k objects from n distinct objects to form a group is given by $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Note that the ordering is **irrelevant** in this case.

Example 2.5 *Computing probability*

Thirteen cards are selected at random from a 52-card deck.

1. How many possible hands are there? Answer: $\binom{52}{13} = 635013559600$

2. What is the probability of getting a hand consisting entirely of spades and clubs with at least one card of each suit?
$$\text{Probability} = \frac{\binom{26}{13} - 2}{\binom{52}{13}} = .000016379$$

3. What is the chance of getting a hand consisting of exactly two suits?
$$\text{Probability} = \frac{\binom{4}{2}\left[\binom{26}{13} - 2\right]}{\binom{52}{13}} = .000098271$$

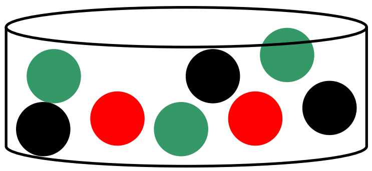
4. What is the chance of getting a hand of 4♠-4♡-3♢-2♣?
$$\text{Probability} = \frac{\binom{13}{4}\binom{13}{4}\binom{13}{3}\binom{13}{2}}{\binom{52}{13}} = 0.017959$$

5. What is the probability of getting a hand with distribution 4-4-3-2? Answer: 0.21551

$$\text{Probability} = \frac{\binom{13}{4}\binom{13}{4}\binom{13}{3}\binom{13}{2}}{\binom{52}{13}}\binom{4}{2}\binom{2}{1}$$

6. What is the chance of getting a hand of K-Q-J if 3 cards are picked at random? Answer: $\frac{\binom{4}{1}\times\binom{4}{1}\times\binom{4}{1}}{\binom{52}{3}}$.

7. Three balls are selected at random without replacement from the jar below. What is the probability that one ball is red and two are black? Answer: $\frac{\binom{2}{1}\binom{3}{2}}{\binom{8}{3}}$.

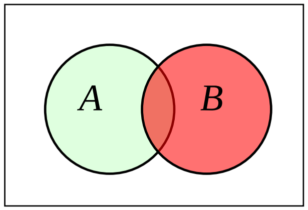


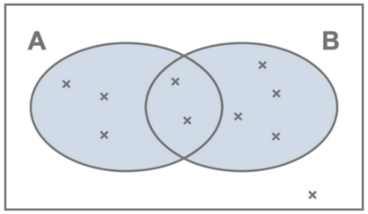
2.4 Conditional probability

Conditioning is powerful for relating two events and provides a useful tool for computing probability.

Definition: Conditional probability of A given B is defined (provided that $P(B) > 0$) as

$$P(A|B) = \frac{P(AB)}{P(B)}$$





Example 2.6 *Conditional probability.*

In a population, 50% of the people are female and 5% of them are female and unemployed. If a randomly chosen person is female, what is the chance that she is unemployed?

$$P(U|F) = \frac{P(F \cap U)}{P(F)} = \frac{.05}{.50} = 10\%.$$

What is $P(U)$? **5%**

Multiplication rule: $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$.

More generally,

$$P(A_1 \cdots A_k) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_k|A_1 \cdots A_{k-1}).$$

Example 2.7 *A birthday problem*

In a class of n students, what is the chance that at least two people have the same birthday?

$$\begin{aligned} p_n &\equiv P(\text{at least two having the same birthday}) \\ &= 1 - P(\text{none of students have the same birthday}) \\ &= 1 - \frac{365}{365} \times \frac{364}{365} \times \cdots \times \frac{365-n+1}{365} \end{aligned}$$

In particular, $p_{23} = 0.5073$, $p_{30} = 0.7063$, $p_{40} = 0.8912$, $p_{50} = 0.97$.

Question: What is the probability that at least one of n students have the same birthday as you?

Let us use simulation to demonstrate the birthday problem:

```
> x = sample(1:365,50,replace=T)      #draw 50 birthdays independently.
> x
 [1] 281 353 140 206 232 175 238 206 240 186 229  68 245  94 182  26 190 361 177
[20] 335 192 301 142 164  74 142  26  95 203  96 266 281 141 190 311 128 267 122
[39] 316 268  53 118 230  23  38 150 362 110 210 145
> table(x)        #if the same birth days, its length is shorter than that of x.
```

```
 23  26  38  53  68  74  94  95  96 110 118 122 128 140 141 142 145 150 164 175
  1   2   1   1   1   1   1   1   1   1   1   1   1   1   1   2   1   1   1   1
177 182 186 190 192 203 206 210 229 230 232 238 240 245 266 267 268 281 301 311
  1   1   1   2   1   1   2   1   1   1   1   1   1   1   1   1   1   2   1   1
316 335 353 361 362
  1   1   1   1   1
```

We now simulate 10000 times and see what is the relative frequency

```
> result = NULL
> for(i in 1:10000)                    #10000 simulations
   {x = sample(1:365,50,replace=T)      #draw 50 birthdays
    result = c(result, (length(table(x)) < length(x)))
   }
> result[1:10]                          #result of the first 10 experiments
 [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
> mean(result)                  #Sample proportion with the same birthdays
[1] 0.9717
```

Example 2.8 *A box contains 10 tickets. Three of them are winning tickets. Pick three at random without replacement.*

1. What is the chance that the first three randomly picked tickets are the winning ones?

   Let W_i be the event that the ith draw is a winning ticket. Then,

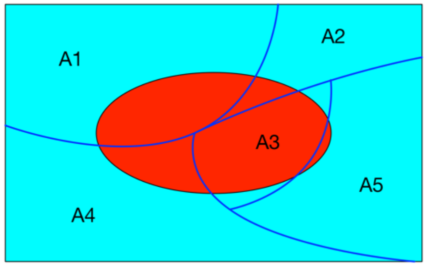
$$\begin{aligned} P(W_1W_2W_3) &= P(W_1)P(W_2|W_1)P(W_3|W_1W_2) \\ &= \frac{3}{10} \times \frac{2}{9} \times \frac{1}{8} = \frac{1}{120}. \end{aligned}$$

2. What is the chance that the second ticket is a winning one?

$$\begin{aligned} P(W_2) &= P(W_1W_2) + P(W_1^cW_2) \\ &= P(W_1)P(W_2|W_1) + P(W_1^c)P(W_2|W_1^c) \\ &= \frac{3}{10} \times \frac{2}{9} + \frac{7}{10} \times \frac{3}{9} = \frac{3}{10}. \end{aligned}$$

Rule of total probability: If $A_1, \cdots, A_k$ are a partition of the sample space, then

$$P(B) = P(A_1)P(B|A_1) + \cdots + P(A_k)P(B|A_k).$$



Example 2.9 *Rule of total probability and Bayes rule*

Lab tests produce positive and negative results. Assume that a lab test has 95% sensitivity and 98% specificity. Assume that the prevalence probability is 1%.

1. What is the chance that lab test is positive for a random person?

$$P(+) = P(D)P(+|D) + P(D^c)P(+|D^c)$$
$$= .01 \times .95 + .99 \times .02 = 2.93\%.$$

2. Pick one person at random and the lab test shows positive result.

What is the chance that the person really has the disease?

$$P(D|+) = \frac{P(D+)}{P(+)} = \frac{P(D)P(+|D)}{P(D)P(+|D) + P(D^c)P(+|D^c)}$$
$$= \frac{.95 \times .01}{01 \times .95 + .99 \times .02} = 32\%.$$

3. What is the probability that, given that the lab test shows a neg-

ative result, the person does not have the disease?

$$P(D^c|-) = \frac{P(D^c-)}{P(-)} = \frac{P(D^c)P(-|D^c)}{P(D)P(-|D) + P(D^c)P(-|D^c)}$$
$$= \frac{\qquad}{\qquad} = 99.95\%.$$

4. If a patient comes to see a doctor and the doctor judges that the probability of having the disease is 30%, what is the chance that the person really has the disease given that the lab test is positive? Answer: 95.3%.

Bayes' Theorem: If $A_1, \cdots, A_k$ are a partition of the sample space, then

$$P(A_j|B) = \frac{P(A_jB)}{P(B)} = \frac{P(A_j)P(B|A_j)}{P(A_1)P(B|A_1) + \cdots + P(A_k)P(B|A_k)}.$$

Basic notion: Compute $A_j|B$ using $B|A_j$, i.e. calculate **posterior** probability using **prior** knowledge.

2.5 Independence

Definition: Two events A and B are indep. if $P(A|B) = P(A)$.

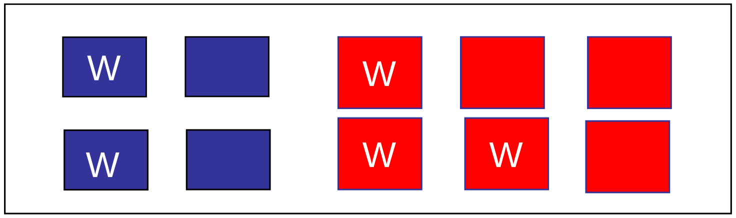
The definition is equivalent to:

(i) $P(AB) = P(A)P(B)$; (ii) $P(A^cB) = P(A^c)P(B)$;

(iii) $P(AB^c) = P(A)P(B^c)$ (iv) $P(A^cB^c) = P(A^c)P(B^c)$.

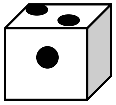
Example 2.10 *Intuition of independence*

A box contains 4 blue tickets and 6 red tickets. Among the blue



tickets, there are two winning ones, and among the red tickets, there are three winning ones. Upon picking a ticket at random from the box you see that it is a red one. What is the chance of getting a winning ticket?

Example 2.11 *Roll a dice.*

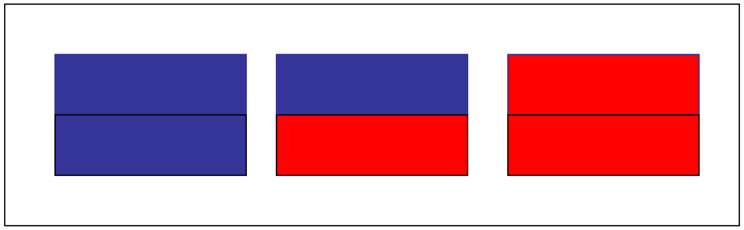


Let $A = \{2, 4, 6\}$, $B = \{1, 2, 3\}$ and $C = \{1, 2, 3, 4\}$. We then have

$$P(A) = 0.5, \qquad P(A|B) = \frac{1}{3}, \qquad P(A|C) = \frac{1}{2}.$$

Namely, A and B are dependent, while A and C are independent

Question: A hat contains 3 index cards: one with both sides blue, one with both sides red, and one with one side blue and the other side red. Assume that the cards are well shuffled. Picking one at random, you see that the front side is blue. What is the chance that the back side is red? **Answer**: 1/3.



Solution: Since the 3 index cards are symmetric in terms of what is "back" and "front", the sample space has in fact 6 outcomes with equal probability (1/6)

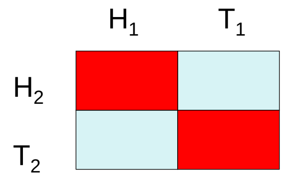
$$\Omega = \{\text{front-back}, \text{front-back}, \text{front-back}, \text{front-back}, \text{front-back}, \text{front-back}\}.$$

Then, by definition of the conditional probability, we have:

$$P(\text{back}|\text{front}) = \frac{P(\text{front} \cap \text{back})}{P(\text{front})} = \frac{1/6}{3/6} = \frac{1}{3},$$

since $P(\text{front}) = P(\text{front} \cap (\text{back} \cup \text{back}) = P[(\text{front} \cap \text{back}) \cup (\text{front} \cap \text{back})] = P(\text{front} \cap \text{back})) + P(\text{front} \cap \text{back})] = 2/6 + 1/6 = 3/6$.

Example 2.12 *Independence of three events*



Suppose that a coin is tossed twice. Let H_1 and T_1 be the events that the first toss is head and tail, respectively. Similar notation applies to the second toss. Let $E = H_1H_2 \bigcup T_1T_2$. Then

- ♠ E, H_1, H_2 are pairwise independent. e.g., $P(EH_1) = P(H_1H_2) = 1/4 = P(E)P(H_1)$.
- ♠ E, H_1, H_2 are not independent, since $P(E|H_1H_2) = 1 \neq P(E) = 0.5$

Definition: Events $A_1, \cdots, A_n$ are mutually independent if for every subset of indices $i_1, \cdots, i_k$ of $\{1, \cdots, n\}$, we have

$$P(A_{i_1} \cdots A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}).$$

(total: 2^n-equations; more than pairwise independence.) This definition can be shown to be equivalent to the conditions

$$P(B_1 \cdots B_n) = P(B_1) \cdots P(B_n), \qquad \text{for all } B_j = A_j \text{ or } A_j^c$$

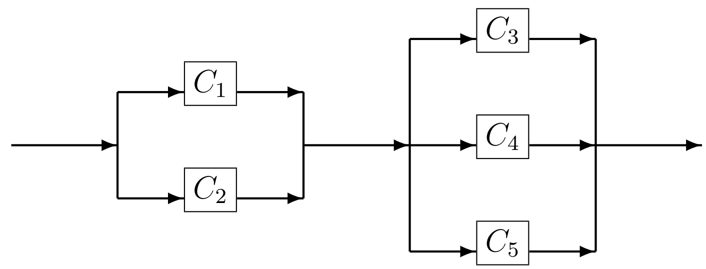
(2^n-equations). Thus, indep is an idealistic assumption in practice.

Example 2.13 *Use of independence*

Suppose that a system consists of 5 indep components $C_1, \cdots C_5$ connected as below. If the probability that each component works properly is 90%, what is the chance that the system works properly?

Let D_1 and D_2 denote the first and second parallel device. Then,

$$P(D_1) = P(C_1 \cup C_2) = 1 - P(C_1^c C_2^c) = 1 - 0.1 * 0.1 = .99.$$



Similarly, $P(D_2) = 1 - 0.1 * 0.1 * 0.1 = 0.999.$

The probability that the system works properly is

$$P(D_1 D_2) = 0.99 \times 0.999 = .9890$$

Note: In the last expression we use the fact that D_1 and D_2 and independent events. Recall that D_1 and D_2 are independent events if and only if D_1^c and D_2^c are independent events. We demonstrate below the latter

$$\begin{aligned} P(D_1^c \cap D_2^c) &= P[(C_1 \cup C_2)^c \cap (C_3 \cup C_4 \cup C_5)^c] \\ &= P(C_1^c \cap C_2^c \cap C_3^c \cap C_4^c \cap C_5^5) \qquad \Leftarrow \text{de Morgan's Law} \\ &= P(C_1^c \cap C_2^c) P(C_3^c \cap C_4^c \cap C_5^5) \qquad \Leftarrow C_1, \ldots, C_5 \text{ are independent then so are } C_1^c, \ldots, C_5^c \\ &= P[(C_1 \cup C_2)^c] P[(C_3 \cup C_4 \cup C_5)^c] \qquad \Leftarrow \text{de Morgan's Law} \\ &= P(D_1^c) P(D_2^c). \end{aligned}$$