

ET 601: Computer Applications for Engineers Event-Based Probability Theory

Asst. Prof. Dr. Prapun Suksompong
prapun@siit.tu.ac.th

December 16, 2013

Contents

1	Review of Set Theory	2
2	Probability Foundations	7
3	Event-based Independence and Conditional Probability	15
3.1	Event-based Conditional Probability	15
3.2	Event-based Independence	25
3.3	Bernoulli Trials	30

1 Review of Set Theory

1.1. Venn diagram is very useful in set theory. It is often used to portray relationships between sets. Many identities can be read out simply by examining Venn diagrams.

Example 1.2. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$

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

Definition 1.4. 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 .

1.5. Basic Set Identities:

- Idempotence: $(A^c)^c = A$
- Commutativity (symmetry):

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

- 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**

- $(A \cup B)^c = A^c \cap B^c$
- $(A \cap B)^c = A^c \cup B^c$

1.6. 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 [1, p. 9] if and only if $A_i \cap A_j = \emptyset$ when $i \neq j$.

Example 1.7. Sets A , B , and C are pairwise disjoint if

1.8. 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 1.9. Given a set S , a collection $\Pi = (A_\alpha : \alpha \in I)$ of subsets¹ 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 1.10 (Slide:maps).

Example 1.11. Let E be the set of students taking ET601

Definition 1.12. 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.

- A *finite* set is a set that has a finite number of elements.
- A set that is not finite is called *infinite*.
- Countable sets:

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

- Empty set and finite sets are automatically countable.
 - 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, \dots
- A **singleton** is a set with exactly one element.
 - Ex. $\{1.5\}$, $\{.8\}$, $\{\pi\}$.
 - *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 .

1.13. We can categorize sets according to their cardinality:

Example 1.14. Examples of countably infinite sets:

- the set $\mathbb{N} = \{1, 2, 3, \dots\}$ 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,

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

Table 1: The terminology of set theory and probability theory

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

Example 1.15. Example of uncountable sets²:

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

Definition 1.16. 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.

1.17. 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.

²For those who are curious, we may use a technique called (Cantor's) diagonal argument (also called the diagonalisation argument, the diagonal slash argument or the diagonal method) to prove that a set is not countable and hence uncountable.

2 Probability Foundations

Constructing the mathematical foundations of probability theory has proven to be a long-lasting process of trial and error. *The approach consisting of defining probabilities as relative frequencies in cases of repeatable experiments leads to an unsatisfactory theory.* The frequency view of probability has a long history that goes back to **Aristotle**. It was not until 1933 that the great Russian mathematician A. N. **Kolmogorov** (1903-1987) laid a satisfactory mathematical foundation of probability theory. He did this by taking a number of axioms as his starting point, as had been done in other fields of mathematics. [7, p 223]

We will try to avoid several technical details³ ⁴ in this class. Therefore, the definition given below is not the “complete” definition. Some parts are modified or omitted to make the definition easier to understand.

³To study formal definition of probability, we start with the **probability space** (Ω, \mathcal{A}, P) . Let Ω be an arbitrary space or set of points ω . Recall that, viewed probabilistically, a subset of Ω is an **event** and an element ω of Ω is a **sample point**. Each event is a collection of outcomes which are elements of the sample space Ω .

The theory of probability focuses on collections of events, called event **σ -algebras**, typically denoted by \mathcal{A} (or \mathcal{F}), that contain all the events of interest (regarding the random experiment \mathcal{E}) to us, and are such that we have knowledge of their likelihood of occurrence. The probability P itself is defined as a number in the range $[0, 1]$ associated with each event in \mathcal{A} .

⁴The class 2^Ω of all subsets can be too large for us to define probability measures with consistency, across all member of the class. (There is no problem when Ω is countable.)

Definition 2.1. Kolmogorov's Axioms for Probability [3]: A **probability measure**⁵ is a real-valued set function⁶ that satisfies

P1 Nonnegativity:

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

P2 Unit normalization:

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

P3 Countable additivity or σ -additivity: For every countable sequence $(A_n)_{n=1}^{\infty}$ of disjoint events,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

- The number $P(A)$ is called the **probability** of the event A
- The entire sample space Ω is called the **sure event** or the **certain event**.
- If an event A satisfies $P(A) = 1$, we say that A is an **almost-sure event**.
- A **support** of P is any set A for which $P(A) = 1$.

From the three axioms⁷, we can derive many more properties of probability measure. These properties are useful for calculating probabilities.

⁵Technically, probability measure is defined on a σ -algebra \mathcal{A} of Ω . The *triple* (Ω, \mathcal{A}, P) is called a **probability measure space**, or simply a **probability space**

⁶A real-valued set function is a function that maps sets to real numbers.

⁷Remark: The axioms do not determine probabilities; the probabilities are assigned based on our knowledge of the system under study. (For example, one approach is to base probability assignments on the simple concept of equally likely outcomes.) The axioms enable us to easily calculate the probabilities of some events from knowledge of the probabilities of other events.

2.2. $P(\emptyset) = 0$.

2.3. Finite additivity⁸: If A_1, \dots, A_n are disjoint events, then

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

Special case when $n = 2$: **Addition rule** (Additivity)

$$\text{If } A \cap B = \emptyset, \text{ then } P(A \cup B) = P(A) + P(B). \quad (1)$$

⁸It is not possible to go backwards and use finite additivity to derive countable additivity (P3).

2.4. The probability of a finite or countable event equals the sum of the probabilities of the outcomes in the event.

(a) In particular, if A is countable, e.g. $A = \{a_1, a_2, \dots\}$, then

$$P(A) = \sum_{n=1}^{\infty} P(\{a_n\}).$$

(b) Similarly, if A is finite, e.g. $A = \{a_1, a_2, \dots, a_{|A|}\}$, then

$$P(A) = \sum_{n=1}^{|A|} P(\{a_n\}).$$

- This greatly simplifies⁹ construction of probability measure.

Remark: Note again that the set A under consideration here is finite or countably infinite. You can not apply the properties above to uncountable set.¹⁰

⁹ Recall that a probability measure P is a set function that assigns number (probability) to all set (event) in \mathcal{A} . When Ω is countable (finite or countably infinite), we may let $\mathcal{A} = 2^\Omega =$ the power set of the sample space. In other words, in this situation, it is possible to assign probability value to all subsets of Ω .

To define P , it seems that we need to specify a large number of values. Recall that to define a function $g(x)$ you usually specify (in words or as a formula) the value of $g(x)$ at all possible x in the domain of g . The same task must be done here because we have a function that maps sets in \mathcal{A} to real numbers (or, more specifically, the interval $[0, 1]$). It seems that we will need to explicitly specify $P(A)$ for each set A in \mathcal{A} . Fortunately, 2.4 implies that we only need to define P for all the singletons (when Ω is countable).

¹⁰In Section ??, we will start talking about (absolutely) continuous random variables. In such setting, we have $P(\{\alpha\}) = 0$ for any α . However, it is possible to have an uncountable set A with $P(A) > 0$. This does not contradict the properties that we discussed in 2.4. If A is finite or countably infinite, we can still write

$$P(A) = \sum_{\alpha \in A} P(\{\alpha\}) = \sum_{\alpha \in A} 0 = 0.$$

For event A that is uncountable, the properties in 2.4 are not enough to evaluate $P(A)$.

Example 2.5. A random experiment can result in one of the outcomes $\{a, b, c, d\}$ with probabilities 0.1, 0.3, 0.5, and 0.1, respectively. Let A denote the event $\{a, b\}$, B the event $\{b, c, d\}$, and C the event $\{d\}$.

- $P(A) =$
- $P(B) =$
- $P(C) =$
- $P(A^c) =$
- $P(A \cap B) =$
- $P(A \cap C) =$

2.6. Monotonicity: If $A \subset B$, then $P(A) \leq P(B)$

Example 2.7. Let A be the event to roll a 6 and B the event to roll an even number. Whenever A occurs, B must also occur. However, B can occur without A occurring if you roll 2 or 4.

2.8. If $A \subset B$, then $P(B \setminus A) = P(B) - P(A)$

2.9. $P(A) \in [0, 1]$.

2.10. $P(A \cap B)$ can not exceed $P(A)$ and $P(B)$. In other words, “the composition of two events is always less probable than (or at most equally probable to) each individual event.”

Example 2.11 (Slides). Experiments by psychologists Kahneman and Tversky.

2.12. Complement Rule:

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

- “The probability that something does not occur can be computed as one minus the probability that it does occur.”
- Named “probability’s Trick Number One” in [2]

2.13. Probability of a union (not necessarily disjoint):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- $P(A \cup B) \leq P(A) + P(B)$.
- Approximation: If $P(A) \gg P(B)$ then we may approximate $P(A \cup B)$ by $P(A)$.

Example 2.14 (Slides). Combining error probabilities from various sources in DNA testing

Example 2.15. In his bestseller *Innumeracy*, John Allen Paulos tells the story of how he once heard a local weatherman claim that there was a 50% chance of rain on Saturday and a 50% chance of rain on Sunday and thus a 100% chance of rain during the weekend. Clearly absurd, but what is the error?

Answer: Faulty use of the addition rule (1)!

If we let A denote the event that it rains on Saturday and B the event that it rains on Sunday, in order to use $P(A \cup B) =$

$P(A) + P(B)$, we must first confirm that A and B cannot occur at the same time ($P(A \cap B) = 0$). More generally, the formula that is always holds regardless of whether $P(A \cap B) = 0$ is given by 2.13:

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

The event “ $A \cap B$ ” describes the case in which it rains both days. To get the probability of rain over the weekend, we now add 50% and 50%, which gives 100%, but we must then subtract the probability that it rains both days. Whatever this is, it is certainly more than 0 so we end up with something less than 100%, just like common sense tells us that we should.

You may wonder what the weatherman would have said if the chances of rain had been 75% each day. [6, p 12]

2.16. Probability of a union of three events:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \end{aligned}$$

2.17. Two bounds:

(a) **Subadditivity or Boole’s Inequality:** If A_1, \dots, A_n are events, not necessarily disjoint, then

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i).$$

(b) **σ -subadditivity or countable subadditivity:** If A_1, A_2, \dots is a sequence of measurable sets, not necessarily disjoint, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

- This formula is known as the *union bound* in engineering.

2.18. If a (finite) collection $\{B_1, B_2, \dots, B_n\}$ is a partition of Ω , then

$$P(A) = \sum_{i=1}^n P(A \cap B_i)$$

Similarly, if a (countable) collection $\{B_1, B_2, \dots\}$ is a partition of Ω , then

$$P(A) = \sum_{i=1}^{\infty} P(A \cap B_i)$$

2.19. Connection to classical probability theory: Consider an experiment with **finite** sample space $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ in which each outcome ω_i is **equally likely**. Note that $n = |\Omega|$.

We must have

$$P(\{\omega_i\}) = \frac{1}{n}, \quad \forall i.$$

Now, given any event finite¹¹ event A , we can apply 2.4 to get

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} \frac{1}{n} = \frac{|A|}{n} = \frac{|A|}{|\Omega|}.$$

We can then say that the probability theory we are working on right now is an extension of the classical probability theory. When the conditons/assumptions of classical probability theory are met, then we get back the defining definition of classical classical probability. The extended part gives us ways to deal with situation where assumptions of classical probability theory are not satisfied.

¹¹In classical probability, the sample space is finite; therefore, any event is also finite.

3 Event-based Independence and Conditional Probability

Example 3.1 (Slides). Diagnostic Tests.

3.1 Event-based Conditional Probability

Definition 3.2. *Conditional Probability*: The conditional probability $P(A|B)$ of event A , given that event $B \neq \emptyset$ occurred, is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (2)$$

- Some ways to say¹² or express the conditional probability, $P(A|B)$, are:
 - the “probability of A , given B ”
 - the “probability of A , knowing B ”
 - the “probability of A happening, knowing B has already occurred”
- Defined only when $P(B) > 0$.
 - If $P(B) = 0$, then it is illogical to speak of $P(A|B)$; that is $P(A|B)$ is not defined.

3.3. *Interpretation*: Sometimes, we refer to $P(A)$ as

- a **a priori probability** , or
- the **prior probability** of A , or
- the **unconditional probability** of A .

It is sometimes useful to interpret $P(A)$ as our knowledge of the occurrence of event A *before* the experiment takes place. Conditional probability $P(A|B)$ is the **updated probability** of the

¹²Note also that although the symbol $P(A|B)$ itself is practical, its phrasing in words can be so unwieldy that in practice, less formal descriptions are used. For example, we refer to “the probability that a tested-positive person has the disease” instead of saying “the conditional probability that a randomly chosen person has the disease given that the test for this person returns positive result.”

event A given that we now know that B occurred (but we still do not know which particular outcome in the set B occurred).

Example 3.4. In diagnostic tests Example 3.1, we learn whether we have the disease from test result. Originally, before taking the test, the probability of having the disease is 0.01%. Being tested positive from the 99%-accurate test *updates* the probability of having the disease to about 1%.

More specifically, let D be the event that the testee has the disease and T_P be the event that the test returns positive result.

- Before taking the test, the probability of having the disease is $P(D) = 0.01\%$.
- Using 99%-accurate test means

$$P(T_P|D) = 0.99 \text{ and } P(T_P^c|D^c) = 0.99.$$

- Our calculation shows that $P(D|T_P) \approx 0.01$.

3.5. “Prelude” to the concept of “independence”:

If the occurrence of B does not give you more information about A , then

$$P(A|B) = P(A) \tag{3}$$

and we say that A and B are *independent*.

- Meaning: “learning that event B has occurred does not change the probability that event A occurs.”

We will soon define “independence” in Section 3.2. Property (3) can be regarded as a “practical” definition for independence. However, there are some “technical” issues¹³ that we need to deal with when we actually define independence.

¹³Here, the statement assume $P(B) > 0$ because it considers $P(A|B)$. The concept of independence to be defined in Section 3.2 will not rely directly on conditional probability and therefore it will include the case where $P(B) = 0$.

3.6. Similar properties to the three probability axioms:

(a) Nonnegativity: $P(A|B) \geq 0$

(b) Unit normalization: $P(\Omega|B) = 1$.

In fact, for any event A such that $B \subset A$, we have $P(A|B) = 1$.

This implies

$$P(\Omega|B) = P(B|B) = 1.$$

(c) Countable additivity: For every countable sequence $(A_n)_{n=1}^{\infty}$ of disjoint events,

$$P\left(\bigcup_{n=1}^{\infty} A_n \mid B\right) = \sum_{n=1}^{\infty} P(A_n|B).$$

- In particular, if $A_1 \perp A_2$,

$$P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B)$$

3.7. More Properties:

- $P(A|\Omega) = P(A)$
- $P(A^c|B) = 1 - P(A|B)$

- $P(A \cap B|B) = P(A|B)$
- $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$.
- $P(A \cap B) \leq P(A|B)$

3.8. When Ω is finite and all outcomes have equal probabilities,

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

This formula can be regarded as the classical version of conditional probability.

Example 3.9. Someone has rolled a fair dice twice. You know that one of the rolls turned up a face value of six. The probability that the other roll turned up a six as well is $\frac{1}{11}$ (not $\frac{1}{6}$). [7, Example 8.1, p. 244]

3.10. Probability of compound events

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

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

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

When we have many sets intersected in the conditioned part, we often use “,” instead of “ \cap ”.

Example 3.11. Most people reason as follows to find the probability of getting two aces when two cards are selected at random from an ordinary deck of cards:

(a) The probability of getting an ace on the first card is $4/52$.

(b) Given that one ace is gone from the deck, the probability of getting an ace on the second card is $3/51$.

(c) The desired probability is therefore

$$\frac{4}{52} \times \frac{3}{51}.$$

[7, p 243]

Question: What about the unconditional probability $P(B)$?

Example 3.12. You know that roughly 5% of all used cars have been flood-damaged and estimate that 80% of such cars will later develop serious engine problems, whereas only 10% of used cars that are not flood-damaged develop the same problems. Of course, no used car dealer worth his salt would let you know whether your car has been flood damaged, so you must resort to probability calculations. What is the probability that your car will later run into trouble?

You might think about this problem in terms of proportions.

If you solved the problem in this way, congratulations. You have just used the law of total probability.

3.13. Total Probability Theorem: If a (finite or infinitely) countable collection of events $\{B_1, B_2, \dots\}$ is a partition of Ω , then

$$P(A) = \sum_i P(A|B_i)P(B_i). \quad (4)$$

This is a formula¹⁴ for computing the probability of an event that can occur in different ways.

3.14. Special case:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

This gives exactly the same calculation as what we discussed in Example 3.12.

¹⁴The tree diagram is useful for helping you understand the process. However, when the number of possible cases is large (many B_i for the partition), drawing the tree diagram may be too time-consuming and therefore you should also learn how to apply the total probability theorem directly without the help of the tree diagram.

Example 3.15. Continue from the “Diagnostic Tests” Example 3.1 and Example 3.4.

$$\begin{aligned} P(T_P) &= P(T_P \cap D) + P(T_P \cap D^c) \\ &= P(T_P | D) P(D) + P(T_P | D^c) P(D^c). \end{aligned}$$

For conciseness, we define

$$p_d = P(D)$$

and

$$p_{TE} = P(T_P | D^c) = P(T_P^c | D).$$

Then,

$$P(T_P) = (1 - p_{TE})p_d + p_{TE}(1 - p_d).$$

3.16. Bayes’ Theorem:

(a) Form 1:

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

(b) Form 2: If a (finite or infinitely) countable collection of events $\{B_1, B_2, \dots\}$ is a partition of Ω , then

$$P(B_k|A) = P(A|B_k) \frac{P(B_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum_i P(A|B_i)P(B_i)}.$$

- Extremely useful for making inferences about phenomena that cannot be observed directly.
- Sometimes, these inferences are described as “reasoning about causes when we observe effects”.

Example 3.17. Continue from the “Disease Testing” Examples 3.1, 3.4, and 3.15:

$$\begin{aligned}
 P(D|T_P) &= \frac{P(D \cap T_P)}{P(T_P)} = \frac{P(T_P|D)P(D)}{P(T_P)} \\
 &= \frac{(1 - p_{TE})p_D}{(1 - p_{TE})p_D + p_{TE}(1 - p_D)}
 \end{aligned}$$

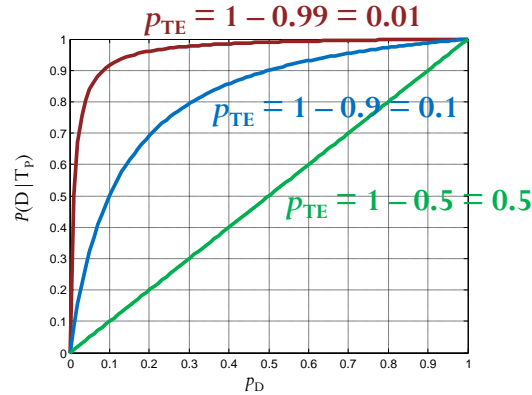


Figure 1: Probability $P(D|T_P)$ that a person will have the disease given that the test result is positive. The conditional probability is evaluated as a function of p_D which tells how common the disease is. The values of test error probability p_{TE} are shown.

Example 3.18. Medical Diagnostic: Because a new medical procedure has been shown to be effective in the early detection of an illness, a medical screening of the population is proposed. The probability that the test correctly identifies someone with the illness as positive is 0.99, and the probability that the test correctly identifies someone without the illness as negative is 0.95. The incidence of the illness in the general population is 0.0001. You take the test, and the result is positive. What is the probability that you have the illness? [5, Ex. 2-37]

Example 3.19. Bayesian networks are used on the Web sites of high-technology manufacturers to allow customers to quickly diagnose problems with products. An oversimplified example is presented here.

A printer manufacturer obtained the following probabilities from a database of test results. Printer failures are associated with three types of problems: hardware, software, and other (such as connectors), with probabilities 0.1, 0.6, and 0.3, respectively. The probability of a printer failure given a hardware problem is 0.9, given a software problem is 0.2, and given any other type of problem is 0.5. If a customer enters the manufacturers Web site to diagnose a printer failure, what is the most likely cause of the problem?

Let the events H , S , and O denote a hardware, software, or other problem, respectively, and let F denote a printer failure.

Example 3.20 (Slides). Prosecutor's Fallacy and the Murder of Nicole Brown

3.21. In practice, here is how we use the total probability theorem and Bayes' theorem:

Usually, we work with a system, which of course has input and output. There can be many possibilities for inputs and there can be many possibilities for output. Normally, for deterministic system, we may have a specification that tells what would be the output given that a specific input is used. Intuitively, we may think of this

as a table of mapping between input and output. For system with random component(s), when a specific input is used, the output is not unique. This means we need conditional probability to describe the output (given an input). Of course, this conditional probability can be different for different inputs.

We will assume that there are many cases that the input can happen. The event that the i th case happens is denoted by B_i . We assume that we consider all possible cases. Therefore, the union of these B_i will automatically be Ω . If we also define the cases so that they do not overlap, then the B_i partitions Ω .

Similarly, there are many cases that the output can happen. The event that the j th case happens is denoted by A_j . We assume that the A_j also partitions Ω .

In this way, the system itself can be described by the conditional probabilities of the form $P(A_j|B_i)$. This replaces the table mentioned above as the specification of the system. Note that even when this information is not available, we can still obtain an approximation of the conditional probability by repeating trials of inputting B_i into the system to find the relative frequency of the output A_j .

Now, when the system is used in actual situation. Different input cases can happen with different probabilities. These are described by the prior probabilities $P(B_i)$. Combining this with the conditional probabilities $P(A_j|B_i)$ above, we can use the total probability theorem to find the probability of occurrence for output and, even more importantly, for someone who cannot directly observe the input, Bayes' theorem can be used to infer the value (or the probability) of the input from the observed output of the system.

In particular, total probability theorem deals with the calculation of the output probabilities $P(A_j)$:

$$P(A_j) = \sum_i P(A_j \cap B_i) = \sum_i P(A_j|B_i) P(B_i).$$

Bayes' theorem calculates the probability that B_k was the input event when the observer can only observe the output of the system

and the observed value of the output is A_j :

$$P(B_k | A_j) = \frac{P(A_j \cap B_k)}{P(A_j)} = \frac{P(A_j | B_k) P(B_k)}{\sum_i P(A_j | B_i) P(B_i)}.$$

Example 3.22. In the early 1990s, a leading Swedish tabloid tried to create an uproar with the headline “Your ticket is thrown away!”. This was in reference to the popular Swedish TV show “Bingolotto” where people bought lottery tickets and mailed them to the show. The host then, in live broadcast, drew one ticket from a large mailbag and announced a winner. Some observant reporter noticed that the bag contained only a small fraction of the hundreds of thousands tickets that were mailed. Thus the conclusion: Your ticket has most likely been thrown away!

Let us solve this quickly. Just to have some numbers, let us say that there are a total of $N = 100,000$ tickets and that $n = 1,000$ of them are chosen at random to be in the final drawing. If the drawing was from all tickets, your chance to win would be $1/N = 1/100,000$. The way it is actually done, you need to both survive the first drawing to get your ticket into the bag and then get your ticket drawn from the bag. The probability to get your entry into the bag is $n/N = 1,000/100,000$. The conditional probability to be drawn from the bag, given that your entry is in it, is $1/n = 1/1,000$. Multiply to get $1/N = 1/100,000$ once more. There were no riots in the streets. [6, p 22]

3.23. Chain rule of conditional probability [1, p 58]:

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

3.2 Event-based Independence

Plenty of random things happen in the world all the time, most of which have nothing to do with one another. If you toss a coin and I roll a dice, the probability that you get heads is $1/2$ regardless of the outcome of my dice. Events that are unrelated to each other in this way are called *independent*.

Definition 3.24. Two events A , B are called (statistically¹⁵) *independent* if

$$P(A \cap B) = P(A)P(B) \quad (5)$$

- Notation: $A \perp\!\!\!\perp B$
- Read “ A and B are independent” or “ A is independent of B ”
- We call (5) the *multiplication rule* for probabilities.
- If two events are not independent, they are *dependent*. Intuitively, if two events are dependent, the probability of one changes with the knowledge of whether the other has occurred.

3.25. Intuition: Again, here is how you should think about independent events: “If one event has occurred, the probability of the other does not change.”

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B). \quad (6)$$

In other words, “the unconditional and the conditional probabilities are the same”. We can almost use (6) as the definitions for independence. This is what we mentioned in 3.5. However, we use (5) instead because it (1) also works with events whose probabilities are zero and (2) also has clear symmetry in the expression (so that $A \perp\!\!\!\perp B$ and $B \perp\!\!\!\perp A$ can clearly be seen as the same). In fact, in 3.29, we show how (6) can be used to define independence with extra condition that deals with the case when zero probability is involved.

¹⁵Sometimes our definition for independence above does not agree with the everyday-language use of the word “independence”. Hence, many authors use the term “statistically independence” to distinguish it from other definitions.

Example 3.26. [8, Ex. 5.4] Which of the following pairs of events are independent?

(a) The card is a club, and the card is black.

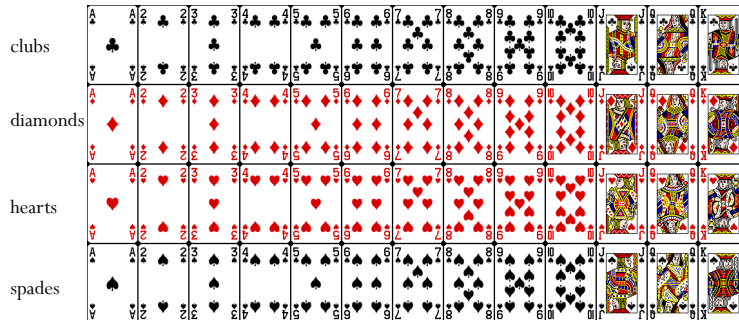


Figure 2: A Deck of Cards

(b) The card is a king, and the card is black.

3.27. An event with probability 0 or 1 is independent of any event (including itself).

- In particular, \emptyset and Ω are independent of any events.

3.28. An event A is independent of itself if and only if $P(A)$ is 0 or 1.

3.29. Two events A, B with positive probabilities are independent if and only if $P(B|A) = P(B)$, which is equivalent to $P(A|B) = P(A)$.

When A and/or B has zero probability, A and B are automatically independent.

3.30. When A and B have nonzero probabilities, the following statements are equivalent:

3.31. The following four statements are equivalent:

$$A \perp\!\!\!\perp B, \quad A \perp\!\!\!\perp B^c, \quad A^c \perp\!\!\!\perp B, \quad A^c \perp\!\!\!\perp B^c.$$

Example 3.32. If $P(A|B) = 0.4$, $P(B) = 0.8$, and $P(A) = 0.5$, are the events A and B independent? [5]

3.33. Keep in mind that **independent and disjoint** are *not synonyms*. In some contexts these words can have similar meanings, but this is not the case in probability.

- If two events cannot occur at the same time (they are disjoint), are they independent? At first you might think so. After all, they have nothing to do with each other, right? Wrong! They have a lot to do with each other. If one has occurred, we know for certain that the other cannot occur. [6, p 12]
- To check whether A and B are disjoint, you only need to look at the sets themselves and see whether they have shared element(s). This can be answered without knowing probabilities.

To check whether A and B are independent, you need to look at the probabilities $P(A)$, $P(B)$, and $P(A \cap B)$.

- Reminder: If events A and B are disjoint, you calculate the probability of the union $A \cup B$ by adding the probabilities of A and B . For independent events A and B you calculate the probability of the intersection $A \cap B$ by multiplying the probabilities of A and B .

- The two statements $A \perp B$ and $A \perp\!\!\!\perp B$ can occur simultaneously only when $P(A) = 0$ and/or $P(B) = 0$.

- Reverse is not true in general.

Example 3.34. Experiment of flipping a fair coin twice. $\Omega = \{HH, HT, TH, TT\}$. Define event A to be the event that the first flip gives a H; that is $A = \{HH, HT\}$. Event B is the event that the second flip gives a H; that is $B = \{HH, TH\}$. Note that even though the events A and B are not disjoint, they are independent.

Example 3.35 (Slides). *Prosecutor’s fallacy* [4, 118–119] and [6, 22–23].

Definition 3.36. Three events A_1, A_2, A_3 are independent if and only if

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

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

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

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

Remarks:

- (a) When the first three equations hold, we say that the three events are *pairwise independent*.
- (b) We may use the term “mutually independence” to further emphasize that we have “independence” instead of “pairwise independence”.

Definition 3.37. The events A_1, A_2, \dots, A_n are *independent* if and only if for any subcollection $A_{i_1}, A_{i_2}, \dots, A_{i_k}$,

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \times P(A_{i_2}) \times \dots \times P(A_{i_k}).$$

- Note that part of the requirement is that

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \times P(A_2) \times \dots \times P(A_n).$$

Therefore, if someone tells us that the events A_1, A_2, \dots, A_n are independent, then one of the properties that we can conclude is that

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \times P(A_2) \times \dots \times P(A_n).$$

- Equivalently, this is the same as the requirement that

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j) \quad \forall J \subset [n] \text{ and } |J| \geq 2$$

- Note that the case when $j = 1$ automatically holds. The case when $j = 0$ can be regarded as the \emptyset event case, which is also trivially true.

3.38. Four events A, B, C, D are pairwise independent if and only if they satisfy the following six conditions:

$$\begin{aligned} P(A \cap B) &= P(A)P(B), \\ P(A \cap C) &= P(A)P(C), \\ P(A \cap D) &= P(A)P(D), \\ P(B \cap C) &= P(B)P(C), \\ P(B \cap D) &= P(B)P(D), \text{ and} \\ P(C \cap D) &= P(C)P(D). \end{aligned}$$

They are independent if and only if they are pairwise independent (satisfy the six conditions above) and also satisfy the following five more conditions:

$$\begin{aligned} P(B \cap C \cap D) &= P(B)P(C)P(D), \\ P(A \cap C \cap D) &= P(A)P(C)P(D), \\ P(A \cap B \cap D) &= P(A)P(B)P(D), \\ P(A \cap B \cap C) &= P(A)P(B)P(C), \text{ and} \\ P(A \cap B \cap C \cap D) &= P(A)P(B)P(C)P(D). \end{aligned}$$

3.3 Bernoulli Trials

Example 3.39. Consider the following random experiments

- (a) Flip a coin 10 times. We are interested in the number of heads obtained.
- (b) Of all bits transmitted through a digital transmission channel, 10% are received in error. We are interested in the number of bits in error in the next five bits transmitted.
- (c) A multiple-choice test contains 10 questions, each with four choices, and you guess at each question. We are interested in the number of questions answered correctly.

These examples illustrate that a general probability model that includes these experiments as particular cases would be very useful.

Example 3.40. Each of the random experiments in Example 3.39 can be thought of as consisting of a series of repeated, random trials. In all cases, we are interested in the number of trials that meet a specified criterion. The outcome from each trial either meets the criterion or it does not; consequently, each trial can be summarized as resulting in either a success or a failure.

Definition 3.41. A *Bernoulli trial* involves performing an experiment once and noting whether a particular event A occurs.

The outcome of the Bernoulli trial is said to be

- (a) a “success” if A occurs and
- (b) a “failure” otherwise.

We may view the outcome of a single Bernoulli trial as the outcome of a toss of an unfair coin for which the probability of heads (success) is $p = P(A)$ and the probability of tails (failure) is $1 - p$.

- The labeling (“success” and “failure”) is not meant to be literal and sometimes has nothing to do with the everyday meaning of the words. We can just as well use A and B or 0 and 1 .

Example 3.42. Examples of Bernoulli trials: Flipping a coin, deciding to vote for candidate A or candidate B, giving birth to a boy or girl, buying or not buying a product, being cured or not being cured, even dying or living are examples of Bernoulli trials.

- Actions that have multiple outcomes can also be modeled as Bernoulli trials if the question you are asking can be phrased in a way that has a yes or no answer, such as “Did the dice land on the number 4?” or “Is there any ice left on the North Pole?”

Definition 3.43. (Independent) *Bernoulli Trials* = a Bernoulli trial that is repeated many times.

- (a) It is usually assumed that the trials are independent. This implies that the outcome from one trial has no effect on the outcome to be obtained from any other trial.
- (b) Furthermore, it is often reasonable to assume that the probability of a success in each trial is constant.

An outcome of the complete experiment is a sequence of successes and failures which can be denoted by a *sequence of ones and zeroes*.

Example 3.44. If we toss *unfair coin* n times, we obtain the space $\Omega = \{H, T\}^n$ consisting of 2^n elements of the form $(\omega_1, \omega_2, \dots, \omega_n)$ where $\omega_i = H$ or T .

Example 3.45. What is the probability of two failures and three successes in five Bernoulli trials with success probability p .

We observe that the outcomes with three successes in five trials are 11100, 11010, 11001, 10110, 10101, 10011, 01110, 01101, 01011, and 00111. We note that the probability of each outcome is a product of five probabilities, each related to one Bernoulli trial. In outcomes with three successes, three of the probabilities are p and the other two are $1 - p$. Therefore, each outcome with three successes has probability $(1 - p)^2 p^3$. There are 10 of them. Hence, the total probability is $10(1 - p)^2 p^3$

3.46. The probability of exactly n_1 success in $n = n_0 + n_1$ bernoulli trials is

$$\binom{n}{n_1} (1-p)^{n-n_1} p^{n_1} = \binom{n}{n_0} (1-p)^{n_0} p^{n-n_0}.$$

Example 3.47. At least one occurrence of a 1-in- n -chance event in n repeated trials:

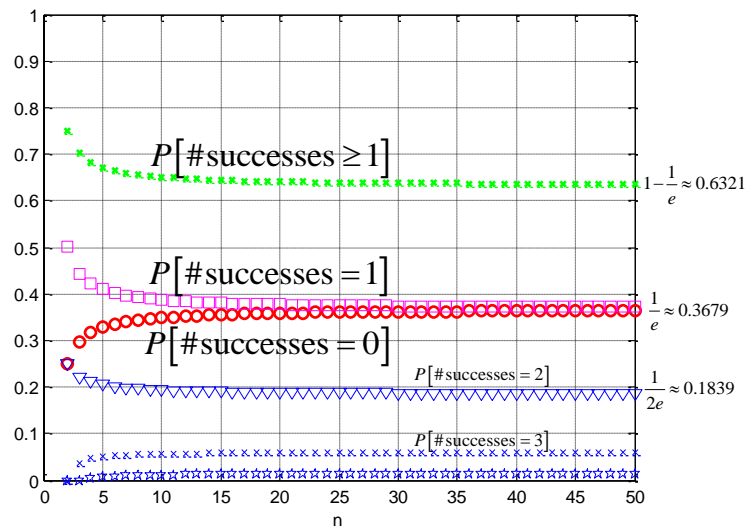


Figure 3: A 1-in- n -chance event in n repeated Bernoulli trials

References

- [1] John A. Gubner. *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge University Press, 2006. 1.6, 3.23
- [2] John Haigh. *Taking Chances: Winning with Probability*. Oxford University Press, 2003. 2.12
- [3] A.N. Kolmogorov. *The Foundations of Probability*. 1933. 2.1
- [4] Leonard Mlodinow. *The Drunkard's Walk: How Randomness Rules Our Lives*. Pantheon; 8th Printing edition, 2008. 3.35
- [5] Douglas C. Montgomery and George C. Runger. *Applied Statistics and Probability for Engineers*. Wiley, 2010. 3.18, 3.32
- [6] Peter Olofsson. *Probabilities: The Little Numbers That Rule Our Lives*. Wiley, 2006. 2.15, 3.22, 3.33, 3.35
- [7] Henk Tijms. *Understanding Probability: Chance Rules in Everyday Life*. Cambridge University Press, 2 edition, August 2007. 2, 3.9, 3
- [8] Rodger E. Ziemer and William H. Tranter. *Principles of Communications*. John Wiley & Sons Ltd, 2010. 3.26