Conditional Probability, Total Probability Theorem and Bayes' Rule

Berlin Chen
Department of Computer Science & Information Engineering
National Taiwan Normal University

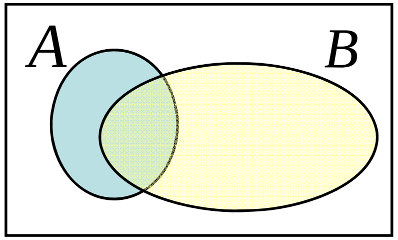
Reference:

- D. P. Bertsekas, J. N. Tsitsiklis, *Introduction to Probability* , Sections 1.3-1.4

Conditional Probability (1/2)

- Conditional probability provides us with a way to reason about the outcome of an experiment, based on partial information
  - Suppose that the outcome is within some given event B, we wish to quantify the likelihood that the outcome also belongs some other given event A
  - Using a new probability law, we have the **conditional probability of A given B**, denoted by $\mathbf{P}(A|B)$, which is defined as:

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



    - If $\mathbf{P}(B)$ has zero probability, $\mathbf{P}(A|B)$ is undefined
    - We can think of $\mathbf{P}(A|B)$ as out of the total probability of the elements of B, the fraction that is assigned to possible outcomes that also belong to A

Conditional Probability (2/2)

- When all outcomes of the experiment are equally likely, the conditional probability also can be defined as

$$\mathbf{P}(A|B) = \frac{\text{number of elements of } A \cap B}{\text{number of elements of } B}$$

- Some examples having to do with conditional probability
  1. 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?
  2. In a word guessing game, the first letter of the word is a "t". What is the likelihood that the second letter is an "h"?
  3. How likely is it that a person has a disease given that a medical test was negative?
  4. A spot shows up on a radar screen. How likely is it that it corresponds to an aircraft?

Conditional Probabilities Satisfy the Three Axioms

- Nonnegative:

$$\mathbf{P}(A|B) \geq 0$$

- Normalization:

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

- Additivity: If A_1 and A_2 are two disjoint events

$$\mathbf{P}(A_1 \cup A_2 | B) = \frac{\mathbf{P}((A_1 \cup A_2) \cap B)}{\mathbf{P}(B)}$$

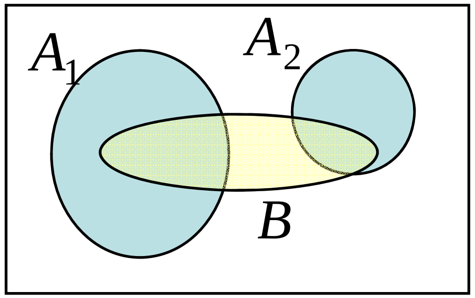
distributive

$$= \frac{\mathbf{P}((A_1 \cap B) \cup (A_2 \cap B))}{\mathbf{P}(B)}$$

disjoint sets

$$= \frac{\mathbf{P}(A_1 \cap B) + \mathbf{P}(A_2 \cap B)}{\mathbf{P}(B)}$$

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



Conditional Probabilities Satisfy General Probability Laws

- Properties probability laws
  - $\boldsymbol{P}(A_1 \cup A_2 | B) \leq \boldsymbol{P}(A_1 | B) + \boldsymbol{P}(A_2 | B)$
  - $\boldsymbol{P}(A_1 \cup A_2 | B) = \boldsymbol{P}(A_1 | B) + \boldsymbol{P}(A_2 | B) - \boldsymbol{P}(A_1 \cap A_2 | B)$
  - …

> Conditional probabilities can also be viewed as a probability law on a new universe B, because all of the conditional probability is concentrated on B.

Simple Examples using Conditional Probabilities (1/3)

Example 1.6. We toss a fair coin three successive times. We wish to find the conditional probability $\mathbf{P}(A \mid B)$ when A and B are the events

$$A = \{\text{more heads than tails come up}\}, \qquad B = \{\text{1st toss is a head}\}.$$

The sample space consists of eight sequences,

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},$$

which we assume to be equally likely. The event B consists of the four elements HHH, HHT, HTH, HTT, so its probability is

$$\mathbf{P}(B) = \frac{4}{8}.$$

The event $A \cap B$ consists of the three elements outcomes HHH, HHT, HTH, so its probability is

$$\mathbf{P}(A \cap B) = \frac{3}{8}.$$

Thus, the conditional probability $\mathbf{P}(A \mid B)$ is

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

Because all possible outcomes are equally likely here, we can also compute $\mathbf{P}(A \mid B)$ using a shortcut. We can bypass the calculation of $\mathbf{P}(B)$ and $\mathbf{P}(A \cap B)$, and simply divide the number of elements shared by A and B (which is 3) with the number of elements of B (which is 4), to obtain the same result $3/4$.

Simple Examples using Conditional Probabilities (2/3)

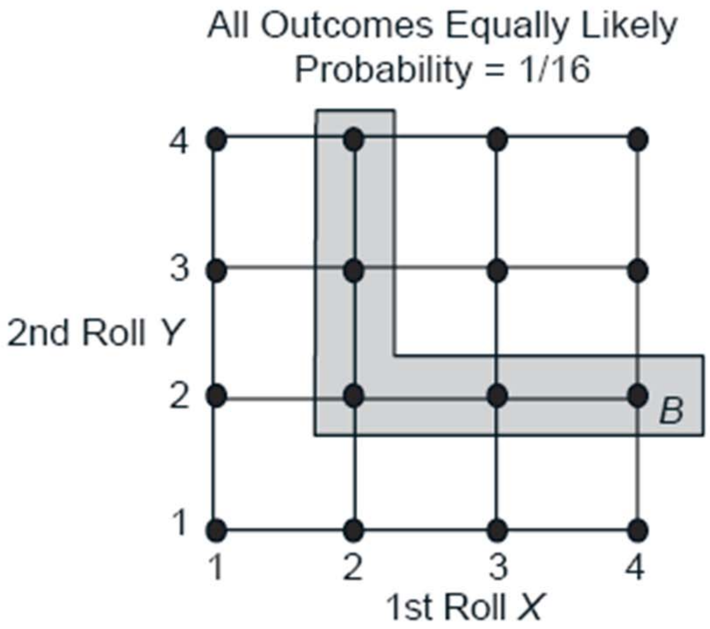


Figure 1.7: Sample space of an experiment involving two rolls of a 4-sided die. (cf. Example 1.7). The conditioning event $B = \{\min(X, Y) = 2\}$ consists of the 5-element shaded set. The set $A = \{\max(X, Y) = m\}$ shares with B two elements if $m = 3$ or $m = 4$, one element if $m = 2$, and no element if $m = 1$. Thus, we have

$$\mathbf{P}\Big(\{\max(X, Y) = m\} \mid B\Big) = \begin{cases} 2/5 & \text{if } m = 3 \text{ or } m = 4, \\ 1/5 & \text{if } m = 2, \\ 0 & \text{if } m = 1. \end{cases}$$

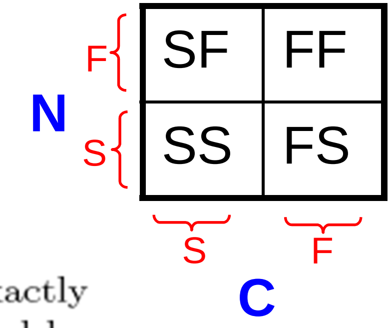
Simple Examples using Conditional Probabilities (3/3)

Example 1.8. A conservative design team, call it C, and an innovative design team, call it N, are asked to separately design a new product within a month. From past experience we know that:

(a) The probability that team C is successful is 2/3.

(b) The probability that team N is successful is 1/2.

(c) The probability that at least one team is successful is 3/4.



If both teams are successful, the design of team N is adopted. Assuming that exactly one successful design is produced, what is the probability that it was designed by team N?

There are four possible outcomes here, corresponding to the four combinations of success and failure of the two teams:

SS: both succeed, FF: both fail,
SF: C succeeds, N fails, FS: C fails, N succeeds.

We are given that the probabilities of these outcomes satisfy

$$\mathbf{P}(SS)+\mathbf{P}(SF)=\frac{2}{3}, \quad \mathbf{P}(SS)+\mathbf{P}(FS)=\frac{1}{2}, \quad \mathbf{P}(SS)+\mathbf{P}(SF)+\mathbf{P}(FS)=\frac{3}{4}.$$

From these relations, together with the normalization equation $\mathbf{P}(SS)+\mathbf{P}(SF)+\mathbf{P}(FS)+\mathbf{P}(FF)=1$, we can obtain the probabilities of all the outcomes:

$$\mathbf{P}(SS)=\frac{5}{12}, \qquad \mathbf{P}(SF)=\frac{1}{4}, \qquad \mathbf{P}(FS)=\frac{1}{12}, \qquad \mathbf{P}(FF)=\frac{1}{4}.$$

The desired conditional probability is

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

Using Conditional Probability for Modeling (1/2)

- It is often natural and convenient to first specify conditional probabilities and then use them to determine unconditional probabilities

- An alternative way to represent the definition of conditional probability

$$\mathbf{P}(A \cap B) = \mathbf{P}(B)\mathbf{P}(A|B)$$

Using Conditional Probability for Modeling (2/2)

Example 1.9. Radar detection. If an aircraft is present in a certain area, a radar correctly registers its presence with probability 0.99. If it is not present, the radar falsely registers an aircraft presence with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of false alarm (a false indication of aircraft presence), and the probability of missed detection (nothing registers, even though an aircraft is present)?

A sequential representation of the sample space is appropriate here, as shown in Fig. 1.8. Let A and B be the events

$$A = \{\text{an aircraft is present}\},$$
$$B = \{\text{the radar registers an aircraft presence}\},$$

and consider also their complements

$$A^c = \{\text{an aircraft is not present}\},$$
$$B^c = \{\text{the radar does not register an aircraft prese}$$

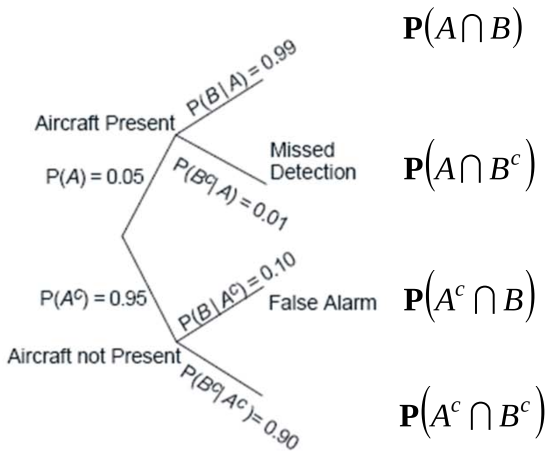


Figure 1.8: Sequential description of the sample space for the radar detection problem in Example 1.9

$$\mathbf{P}(\text{false alarm}) = \mathbf{P}(A^c \cap B) = \mathbf{P}(A^c)\mathbf{P}(B \,|\, A^c) = 0.95 \cdot 0.10 = 0.095,$$
$$\mathbf{P}(\text{missed detection}) = \mathbf{P}(A \cap B^c) = \mathbf{P}(A)\mathbf{P}(B^c \,|\, A) = 0.05 \cdot 0.01 = 0.0005.$$

Multiplication (Chain) Rule

- Assuming that all of the conditioning events have positive probability, we have

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

  – The above formula can be verified by writing

$$\mathbf{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbf{P}(A_1)\frac{\mathbf{P}(A_1 \cap A_2)}{\mathbf{P}(A_1)}\frac{\mathbf{P}(A_1 \cap A_2 \cap A_3)}{\mathbf{P}(A_1 \cap A_2)}\cdots\frac{\mathbf{P}\left(\bigcap_{i=1}^{n} A_i\right)}{\mathbf{P}\left(\bigcap_{i=1}^{n-1} A_i\right)}$$

  – For the case of just two events, the multiplication rule is simply the definition of conditional probability

$$\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2 \mid A_1)$$

Multiplication (Chain) Rule: Examples (1/2)

- **Example 1.10.** Three cards are drawn from an ordinary 52-card deck without replacement (drawn cards are not placed back in the deck). We wish to find the probability that none of the three cards is a "heart".

$$\mathbf{P}(A_i) = \{\text{the } i\text{th card is not a heart}\},\ i = 1,2,3$$

$$\mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(A_1)\mathbf{P}(A_2 | A_1)\mathbf{P}(A_3 | A_1 \cap A_2)$$

$$= \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50}$$

$$\frac{C_3^{39}}{C_3^{52}}?$$

Multiplication (Chain) Rule: Examples (2/2)

- **Example 1.11.** A class consisting of 4 graduate and 12 undergraduate students is randomly divided into 4 groups of 4. What is the probability that each group includes a graduate student?

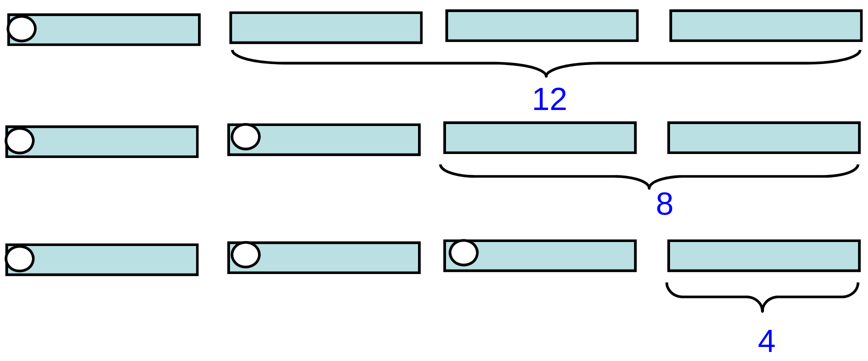
$$A_1 = \{\text{graduate students 1 and 2 are at different groups}\}$$
$$A_2 = \{\text{graduate students 1,2, and 3 are at different groups}\}$$
$$A_3 = \{\text{graduate students 1,2, 3, and 4 are at different groups}\}$$

$$\mathbf{P}(A_3) = \mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(A_1)\mathbf{P}(A_2 | A_1)\mathbf{P}(A_3 | A_1 \cap A_2)$$

$$\mathbf{P}(A_1) = \frac{12}{15}$$

$$\mathbf{P}(A_2 | A_1) = \frac{8}{14}$$

$$\mathbf{P}(A_3 | A_1 \cap A_2) = \frac{4}{13}$$



$$\therefore \mathbf{P}(A_3) = \frac{12}{15} \cdot \frac{8}{14} \cdot \frac{4}{13}$$

Total Probability Theorem (1/2)

- Let $A_1, \cdots, A_n$ 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, we have

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

  - Note that each possible outcome of the experiment (sample space) is included in one and only one of the events $A_1, \cdots, A_n$

Total Probability Theorem (2/2)

Figure 1.13: Visualization and verification of the total probability theorem. The events $A_1, \ldots, A_n$ form a partition of the sample space, so the event B can be decomposed into the disjoint union of its intersections $A_i \cap B$ with the sets A_i, i.e.,

$$B = (A_1 \cap B) \cup \cdots \cup (A_n \cap B).$$

Using the additivity axiom, it follows that

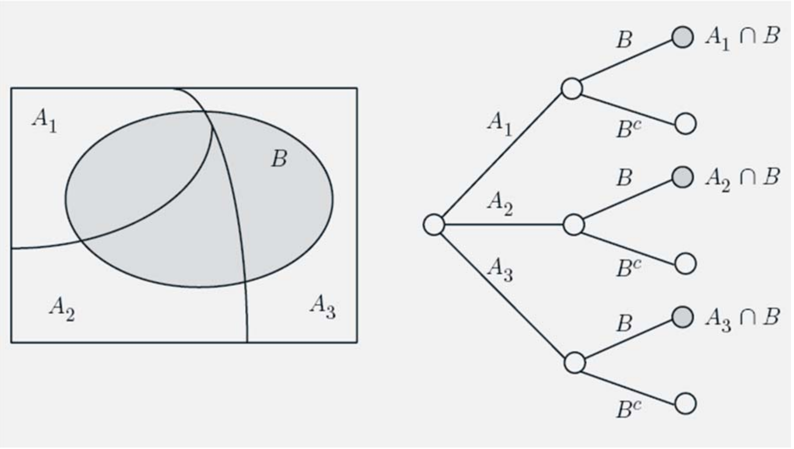
$$\mathbf{P}(B) = \mathbf{P}(A_1 \cap B) + \cdots + \mathbf{P}(A_n \cap B).$$

Since, by the definition of conditional probability, we have

$$\mathbf{P}(A_i \cap B) = \mathbf{P}(A_i)\mathbf{P}(B \mid A_i),$$

the preceding equality yields

$$\mathbf{P}(B) = \mathbf{P}(A_1)\mathbf{P}(B \mid A_1) + \cdots + \mathbf{P}(A_n)\mathbf{P}(B \mid A_n).$$



For an alternative view, consider an equivalent sequential model, as shown on the right. The probability of the leaf $A_i \cap B$ is the product $\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)$ of the probabilities along the path leading to that leaf. The event B consists of the three highlighted leaves and $\mathbf{P}(B)$ is obtained by adding their probabilities.

Some Examples Using Total Probability Theorem (1/3)

Example 1.13. You enter a chess tournament where your probability of winning a game is 0.3 against half the players (call them type 1), 0.4 against a quarter of the players (call them type 2), and 0.5 against the remaining quarter of the players (call them type 3). You play a game against a randomly chosen opponent. What is the probability of winning?

Let A_i be the event of playing with an opponent of type i. We have

$$\mathbf{P}(A_1) = 0.5, \qquad \mathbf{P}(A_2) = 0.25, \qquad \mathbf{P}(A_3) = 0.25.$$

Let also B be the event of winning. We have

$$\mathbf{P}(B \mid A_1) = 0.3, \qquad \mathbf{P}(B \mid A_2) = 0.4, \qquad \mathbf{P}(B \mid A_3) = 0.5.$$

Thus, by the total probability theorem, the probability of winning is

$$\begin{aligned}\mathbf{P}(B) &= \mathbf{P}(A_1)\mathbf{P}(B \mid A_1) + \mathbf{P}(A_2)\mathbf{P}(B \mid A_2) + \mathbf{P}(A_3)\mathbf{P}(B \mid A_3) \\ &= 0.5 \cdot 0.3 + 0.25 \cdot 0.4 + 0.25 \cdot 0.5 \\ &= 0.375.\end{aligned}$$

Some Examples Using Total Probability Theorem (2/3)

Example 1.14. We roll a fair four-sided die. If the result is 1 or 2, we roll once more but otherwise, we stop. What is the probability that the sum total of our rolls is at least 4?

Let A_i be the event that the result of first roll is i, and note that $\mathbf{P}(A_i) = 1/4$ for each i. Let B be the event that the sum total is at least 4. Given the event A_1, the sum total will be at least 4 if the second roll results in 3 or 4, which happens with probability 1/2. Similarly, given the event A_2, the sum total will be at least 4 if the second roll results in 2, 3, or 4, which happens with probability 3/4. Also, given the event A_3, we stop and the sum total remains below 4. Therefore,

$$\mathbf{P}(B \,|\, A_1) = \frac{1}{2}, \qquad \mathbf{P}(B \,|\, A_2) = \frac{3}{4}, \qquad \mathbf{P}(B \,|\, A_3) = 0, \qquad \mathbf{P}(B \,|\, A_4) = 1.$$

(1,3),(1,4) (2,2),(2,3),(2,4) (4)

By the total probability theorem,

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

Some Examples Using Total Probability Theorem (3/3)

- **Example 1.15.** Alice is taking a probability class and 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 a given week, the probability that she will be up-to-date (or behind) in the next week 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 (by default) up-to-date when she starts the class. What is the probability that she is up-to-date after three weeks?

U_i : up-to-date

B_i : behind

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

$$\mathbf{P}(U_2) = \mathbf{P}(U_1)P(U_2|U_1) + \mathbf{P}(B_1)P(U_2|B_1) = \mathbf{P}(U_1)\cdot 0.8 + \mathbf{P}(B_1)\cdot 0.4$$

$$\mathbf{P}(B_2) = \mathbf{P}(B_1)P(B_2|U_1) + \mathbf{P}(B_1)P(B_2|B_1) = \mathbf{P}(U_1)\cdot 0.2 + \mathbf{P}(B_1)\cdot 0.6$$

As we know that $\mathbf{P}(U_1) = 0.8, \quad \mathbf{P}(B_1) = 0.2 \quad (\because \mathbf{P}(U_0) = 1.0)$

$$=> \mathbf{P}(U_2) = 0.8\cdot 0.8 + 0.2\cdot 0.4 = 0.72$$

$$\mathbf{P}(B_2) = 0.8\cdot 0.2 + 0.2\cdot 0.6 = 0.28$$

$$\therefore \quad \mathbf{P}(U_3) = 0.72\cdot 0.8 + 0.28\cdot 0.4 = 0.688$$

Recursion formulea

$$\mathbf{P}(U_{i+1}) = \mathbf{P}(U_i)\cdot 0.8 + \mathbf{P}(B_i)\cdot 0.4$$

$$\mathbf{P}(B_{i+1}) = \mathbf{P}(U_i)\cdot 0.2 + \mathbf{P}(B_i)\cdot 0.6$$

$$\mathbf{P}(U_1) = 0.8, \quad \mathbf{P}(B_1) = 0.2$$

Bayes' Rule

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

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

Multiplication rule

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

Total probability theorem

$$= \frac{\boldsymbol{P}(A_i)\boldsymbol{P}(B|A_i)}{\sum_{k=1}^{n} \boldsymbol{P}(A_k)\boldsymbol{P}(B|A_k)}$$

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

Inference Using Bayes' Rule (1/2)

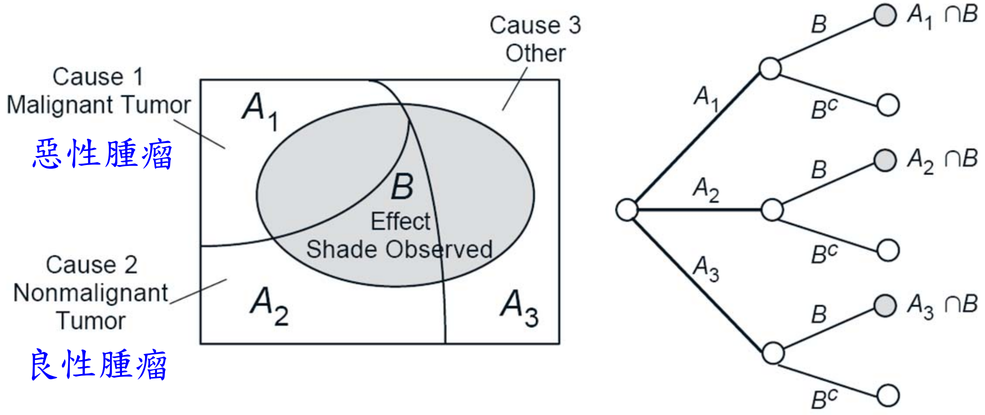


Figure 1.14: An example of the inference context that is implicit in Bayes' rule. We observe a shade in a person's X-ray (this is event B, the "effect") and we want to estimate the likelihood of three mutually exclusive and collectively exhaustive potential causes: cause 1 (event A_1) is that there is a malignant tumor, cause 2 (event A_2) is that there is a nonmalignant tumor, and cause 3 (event A_3) corresponds to reasons other than a tumor. We assume that we know the probabilities $\mathbf{P}(A_i)$ and $\mathbf{P}(B \mid A_i)$, $i = 1, 2, 3$. Given that we see a shade (event B occurs), Bayes' rule gives the conditional probabilities of the various causes as

$$\mathbf{P}(A_i \mid B) = \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(A_1)\mathbf{P}(B \mid A_1) + \mathbf{P}(A_2)\mathbf{P}(B \mid A_2) + \mathbf{P}(A_3)\mathbf{P}(B \mid A_3)}, \quad i = 1, 2, 3.$$

For an alternative view, consider an equivalent sequential model, as shown on the right. The probability $\mathbf{P}(A_1 \mid B)$ of a malignant tumor is the probability of the first highlighted leaf, which is $\mathbf{P}(A_1 \cap B)$, divided by the total probability of the highlighted leaves, which is $\mathbf{P}(B)$.

Inference Using Bayes' Rule (2/2)

- **Example 1.18. The False-Positive Puzzle.**
  - A test for a certain disease is assumed to be correct 95% of the time: if a person has the disease, the test with are positive with probability 0.95 ($\mathbf{P}(B|A)=0.95$), and if the person does not have the disease, the test results are negative with probability 0.95 ($\mathbf{P}(B^c|A^c)=0.95$). A random person drawn from a certain population has probability 0.001 ($\mathbf{P}(A)=0.001$) of having the disease. Given that the person just tested positive, what is the probability of having the disease ($\mathbf{P}(A|B)$) ?
    - A : the event that the person has a disease
    - B : the event that the test results are positive

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

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

$$=\frac{0.001\cdot 0.95}{0.001\cdot 0.95+0.999\cdot 0.05}=0.0187$$

$$\mathbf{P}(B|A^c)=1-\mathbf{P}(B^c|A^c)=0.05$$

Recitation

- SECTION 1.3 Conditional Probability
  - Problems 11, 14, 15
- SECTION 1.4 Probability Theorem, Bayes' Rule
  - Problems 17, 23, 24, 25

Homework -1

- Chapter 1 : Additional Problems

  (http://www.athenasc.com/CH1-prob-supp.pdf)

  – Problems 2, 8, 13, 18, 24