Probability and Statistics

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

February 3, 2019

Notes

Topics

Notes

Objectives

Objectives:

- Recognize problems for which the Law of Total Probability is applicable
- Use the Law of Total Probability to solve probability problems
- Recognize problems for which Bayes Rule is applicable
- Use Bayes Rule to solve probability problems

Notes

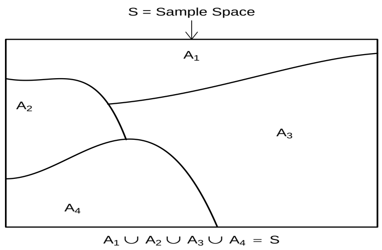
Law of Total Probability (2.4)

- Events $A_1, A_2, \ldots, A_n$ are called a ***partition*** of the sample space $\mathcal{S}$ if

  1. $A_1, A_2, \ldots, A_n$ are mutually exclusive, and
  2. $A_1 \cup A_2 \cup \cdots \cup A_n = \mathcal{S}$.

- If $A_1, A_2, \ldots, A_n$ are a partition of $\mathcal{S}$, then the outcome of the random experiment will satisfy *one and only one* of the A_i's.

Notes



Notes

- The ***Law of Total Probability*** says:

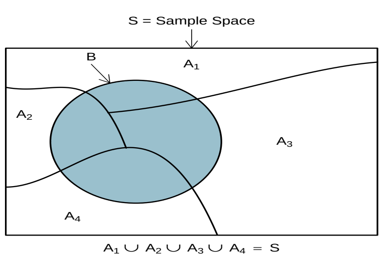
Law of Total Probability: If $A_1, A_2, \ldots, A_n$ are a partition of $\mathcal{S}$, and B is *any* event, then

$$P(B) \;=\; \sum_{i=1}^{n} P(B \cap A_i)$$

which, upon replacing each $P(B \cap A_i)$ by $P(A_i)P(B|A_i)$, can also be written as

$$P(B) \;=\; \sum_{i=1}^{n} P(A_i)P(B|A_i).$$

Notes



Notes

- The **Law of Total Probability** is used to find the **probability** of an event $\boldsymbol{B}$ when:
  1. The probabilities of B occurring under each of several conditions $A_1, A_2, \ldots, A_n$ are known.
  2. The probabilities of each of those conditions being met are known.

Notes

Example

A diagnostic test for a certain disease is **positive 98%** of the time when the disease is **present**, and **(falsely) positive 5%** of the time when the disease is **absent**.

If the disease affects **1 out of every 1,000** people, what's the **probability** that a randomly selected person would test **positive** for the disease?

Notes

To answer this question, we'll apply the **Law of Total Probability**.

We know the **probability** of a **positive** test result (B) under each of two conditions: disease **present** (A_1) and disease **absent** (A_2).

$$P(B \mid A_1) = \mathbf{0.98}$$
$$P(B \mid A_2) = \mathbf{0.05}$$

Notes

We also know the **probability** of each of those two conditions (disease **present**, A_1, and **absent**, A_2).

$$P(A_1) = \frac{\mathbf{1}}{\mathbf{1,000}}$$
$$P(A_2) = \frac{\mathbf{999}}{\mathbf{1,000}}$$

Notes

By the **Law of Total Probability**, the **probability** that a randomly selected person would test **positive** for the disease is

$$\begin{aligned} P(B) &= \sum_{i=1}^{2} P(A_i)P(B|A_i) \\ &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) \\ &= 0.98 \times \frac{1}{1,000} + 0.05 \times \frac{999}{1,000} \\ &= \mathbf{0.0509}. \end{aligned}$$

Notes

Bayes Rule (2.4)

- **Bayes Rule** is used to "flip conditional probabilities around", i.e.

  from $P(B \mid A)$ to $P(A \mid B)$.

  Examples:
  - **Imperfect medical test for a disease** – If a person has a certain disease, there's a 98% chance they'll test positive. If a person tests positive, what's the chance they have the disease?
  - **The "Prosecutor's Fallacy"** – A suspect matching the description of the perpetrator is detained. The chance that an innocent person would match the description is one in a million. What's the chance that a person matching the description is innocent?

Notes

- **Bayes Rule** can be used whenever the Law of Total Probability can be used, i.e. when:
  1. The probabilities of B occurring under each of several conditions $A_1, A_2, \ldots, A_n$ are known.
  2. The probabilities of each of those conditions being met are known.

Notes

- Starting with the definition of **conditional probability**, we get (for any j among $1, 2, \ldots, n$)

$$P(A_j \mid B) = \frac{P(A_j \cap B)}{P(B)}$$

$$= \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)} \quad \begin{array}{l} \leftarrow \text{Mult. Rule} \\ \leftarrow \text{Law of Tot. Prob.} \end{array}$$

which is ***Bayes Rule***:

Bayes' Rule: If $A_1, A_2, \ldots, A_n$ are a partition of $\mathcal{S}$, and B is *any* event, then for each j,

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}.$$

Notes

Example

Suppose again that a disease affects **1 out of every 1,000** people, and a test is **positive 98%** of the time when the disease is **present**, and **(falsely) positive 5%** of the time when the disease is **absent**.

If a randomly selected person tests **positive**, what's the **probability** that the disease is **present**?

Notes

To answer this question, we'll apply **Bayes Rule**.

Recall that the **probabilities** of a **positive** test result ($\boldsymbol{B}$) under each of the conditions disease **present** ($\boldsymbol{A_1}$) and disease **absent** ($\boldsymbol{A_2}$) are

$$P(B \mid A_1) = \mathbf{0.98}$$
$$P(B \mid A_2) = \mathbf{0.05}$$

Notes

Recall also that the **probabilities** of those conditions (disease **present**, $\boldsymbol{A_1}$, and **absent**, $\boldsymbol{A_2}$) are

$$P(A_1) = \frac{\mathbf{1}}{\mathbf{1,000}}$$
$$P(A_2) = \frac{\mathbf{999}}{\mathbf{1,000}}$$

Notes

By the **Bayes Rule**, the **probability** that a person who tests **positive** actually **has the disease** is

$$\begin{aligned} P(A_1|B) &= \frac{P(A_1)P(B|A_1)}{\sum_{i=1}^{2} P(A_i)P(B|A_i)} \\ &= \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)} \\ &= \frac{0.98 \times \frac{1}{1{,}000}}{0.98 \times \frac{1}{1{,}000} + 0.05 \times \frac{999}{1{,}000}} \\ &= \mathbf{0.0192} \end{aligned}$$

i.e. only about a 2% chance!

Notes

- In **Bayes Rule** applications, $\boldsymbol{P(A_j)}$ is called the ***prior probability*** of the event A_j and $\boldsymbol{P(A_j|B)}$ is called the ***posterior probability***.

  The idea is that we use the information that B has occurred to "update" the probability of the event A_j occurring.

Notes