SECTION 4: BAYES' THEOREM

4. Bayes' Theorem (Bayes' Rule)

In probability theory and statistics, Bayes' theorem (alternatively Bayes' law or Bayes' rule) describes the probability of an event, based on conditions that might be related to the event. For example, if cancer is related to age, then, using Bayes' theorem, a person's age can be used to more accurately assess the probability that they have cancer.

One of the many applications of Bayes' theorem is Bayesian inference, a particular approach to statistical inference. When applied, the probabilities involved in Bayes' theorem may have different probability interpretations. With the Bayesian probability interpretation the theorem expresses how a subjective degree of belief should rationally change to account for evidence. Bayesian inference is fundamental to Bayesian statistics.

In many situations, the outcome of an experiment can be observed conditionally to other events. In other words, the outcome of an experiment could not be observed directly. For example, there are n_1 white balls and m_1 black balls are in the first urn, and there are n_2 white balls and m_2 black balls are in the second urn. What is the probability of a black ball drawn?

Theorem 4.1. If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S and $P(B_i) \neq 0$, for i=1,2,…,k, then for any event of A in S

$$P(A) = \sum_{i=1}^{k} P(B_i) P\left(A \middle| B_i\right)$$

This theorem is called the **rule of total probability** or the **rule of elimination**.

Example 4.1. The completion of construction job may be delayed because of a strike. The probabilities are 0.60 that there will be a strike, 0.85 that the construction job will be completed on time if there is no strike, and 0.35 that the construction job will be completed on time if there is a strike. What is the probability that the construction job will be completed on time? (Miller and Miller, 2004, page 48)

Solution:

A is the event that the construction job will be completed on time.
B is the event that there will be strike. $P(B)$=0.60
$P\left(A \middle| B\right) = 0.35$ and $P\left(A \middle| B'\right) = 0.85$

$$\begin{aligned} P(A) &= P\left[(A \cap B) \cup (A \cap B')\right] \\ &= P(A \cap B) + P(A \cap B') \\ &= P(B) P\left(A \middle| B\right) + P(B') P\left(A \middle| B'\right) \\ &= (0.60)(0.35) + (0.40)(0.85) = 0.55 \end{aligned}$$

Example 4.2. The members of consulting firm rent cars from three rental agencies: 60% from agency I, 30% from agency II, and 10% from agency III. If 9% of the cars from agency I need a tune-up (ayar vermek), 20% of the cars from agency II need a tune-up, and 6 % of the cars from agency III need a tune-up, what is the probability that a rental car delivered to the firm will need a tune-up?

Solution:

$P(I) = 0.60, \quad P\left(T \middle| I\right) = 0.09$

$P(II) = 0.30, \quad P\left(T \middle| II\right) = 0.20$

$P(III) = 0.10, \quad P\left(T \middle| III\right) = 0.06$

$$\begin{aligned} P(T) &= P(I)P\left(T \middle| I\right) + P(II)P\left(T \middle| II\right) + P(III)P\left(T \middle| III\right) \\ &= (0.60)(0.09) + (0.30)(0.20) + (0.10)(0.06) = 0.12 \end{aligned}$$

With 12% of cars delivered to the firm will need a tune-up.

Theorem 4.2. If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S and $P(B_i)\neq 0$, for i=1,2,…,k, then for any event of A in S such that $P(A)\neq 0$

$$P\left(B_i \middle| A\right) = \frac{P(B_i)P\left(A \middle| B_i\right)}{\sum_{i=1}^{k} P(B_i)P\left(A \middle| B_i\right)}, \qquad for\; i = 1, 2, \ldots, \mathrm{k}$$

The theorem is called the **Bayes' Theorem**. Where the unconditional probabilities $P(B_i)$ are called prior probabilities and the conditional probabilities $P\left(A \middle| B_i\right)$ are called likelihoods, hence the probabilities $P\left(B_i \middle| A\right)$ are called posterior probabilities.

Example 4.3. Suppose that a laboratory test on a blood sample yields one of two results, positive, negative. It is found that 95% of people with a particular disease produce a positive result. But 2% of people without disease will also produce a positive result (a false positive). Suppose that 1% of the population actually has the disease. What is the probability that person chosen at random from the population will have the disease, given that person's blood yields a positive result?

Solution:

$P\left(+ \middle| D\right) = 0.95, \quad P\left(+ \middle| \bar{D}\right) = 0.02, \quad P(D) = 0.01, \quad P(\bar{D}) = 0.99$

$$\begin{aligned} P\left(D \middle| +\right) &= \frac{P\left(+ \middle| D\right)P(D)}{P\left(+ \middle| D\right)P(D)P\left(+ \middle| \bar{D}\right)P(\bar{D})} \\ &= \frac{(0.95)(0.01)}{(0.95)(0.01) + (0.02)(0.99)} = \frac{95}{293} \cong 32\% \end{aligned}$$

	Test result		total
	Positive(+)	**Negative(-)**	
Have a disease	100×95%=95	100×5%=5	100
Does not have disease	9900×2%=198	9900×98%=9702	9900
total	293	9707	10000

Example 4.4. A manufacturing process produces integrated circuit chips. Over the long run the fraction of bad chips produced by the process is round 20%. Thoroughly testing a chip to determine whether it is good or bad is rather expensive, so a cheap test is tried. All good chips will pass the cheap test, but so will 10% of bad chips.

a) Given a chip passes the cheap test, what is the probability that it is a good chip?
b) If a company using this manufacturing process sell all chips which pass the cheap test, over the long run what percentage of chips sold will be bad?

Solution:
Bad={a chip produced being a bad }

$P\left(+\middle|Good\right)=1,\quad P\left(+\middle|Bad\right)=0.10,\quad P(Bad)=0.20,\quad P(Good)=0.80$

$P(Bad)=1-P(Good)$

a)

$$P\left(Good\middle|+\right)=\frac{P\left(+\middle|Good\right)P(Good)}{P\left(+\middle|Good\right)P(Good)+P\left(+\middle|Bad\right)P(Bad)}$$

$$=\frac{(1)(0.80)}{(1)(0.80)+(0.10)(0.20)}=\frac{80}{82}\cong 98\%$$

b) $P\left(Bad\middle|+\right)=1-P\left(Good\middle|+\right)=1-0.98=2\%$

For example, suppose that if 1 000 000 chips were sold, %2 of them would be bad chips. That is, 20 000 bad chips were sold.

Exercises:

1) The probability that a one-car-accident is due to faulty brakes is 0.04, the probability that a one-car accident is correctly attributed to faulty brakes is 0.82, the probability that a one-car accident is incorrectly attributed to faulty brakes is 0.03. what is the probability that

   a) a one car accident will be attributed to faulty brakes;
   b) a one car accident attributed to faulty brakes was actually due to faulty brakes.

Solution:

a)

F={ faulty break } P(F)=0.04

A={ attributed to faulty brakes }

$P\left(A\middle|F\right)=0.82$ corresponds to the probability of the event about that a one-car accident is correctly attributed to faulty brakes.

$P\left(A\middle|\bar{F}\right)=0.03$ corresponds to the probability of the event about that a one-car accident is incorrectly attributed to faulty brakes.

$$\begin{aligned} P\left(A\right) &= P\left(A\middle|F\right)P(F)+P\left(A\middle|\bar{F}\right)P(\bar{F}) \\ &= (0.82)(0.04)+(0.03)(0.96) \\ &= 0.0616 \end{aligned}$$

b)

$$\begin{aligned} P\left(F\middle|A\right) &= \frac{P\left(A\middle|F\right)P(F)}{P\left(A\middle|F\right)P(F)+P\left(A\middle|\bar{F}\right)P(\bar{F})} \\ &= \frac{(0.82)(0.04)}{0.0616}=\frac{0.0328}{0.0616}\cong 0.5325 \end{aligned}$$

2) In a certain community, 8 percent of all adults over 50 have diabetes. If a health service in this community correctly diagnoses 95 percent of all persons with diabetes as having the disease and incorrectly diagnoses 2 percent of all persons without diabetes as having the disease, find probabilities that

   a) the community health service will diagnose an adult over 50 as having diabetes;
   b) a person over 50 diagnosed by the health service as having diabetes actually has the disease.

Solution:

I={having diabetes disease(illness)}, D={diagnosed diabetes }

$$P\left(D\middle|I\right)=0.95,\quad P\left(D\middle|\bar{I}\right)=0.02$$

a)

$$\begin{aligned} P(D) &= P(D\cap I)+P(D\cap \bar{I}) \\ &= P\left(D\middle|I\right)P(I)+P\left(D\middle|\bar{I}\right)P(\bar{I}) \\ &= (0.95)(0.08)+(0.02)(0.92)=0.0944 \end{aligned}$$

b)

$$\begin{aligned} P\left(I\middle|D\right) &= \frac{P\left(D\middle|I\right)P(I)}{P(D)} \\ &= \frac{(0.95)(0.08)}{0.0944}\cong 0.8051 \end{aligned}$$

3) An explosion at a construction site could have occurred as the result of **static electricity**, **malfunctioning of equipment**, **carelessness**, or **sabotage**. Interviews with construction engineers analyzing the risks involved led to the estimates that such an explosion would occur with probability 0.25 as a result of static electricity, 0.20 as a result of malfunctioning of equipment, 0.40 as result of carelessness, and 0.75 as a result of sabotage. It is also felt that the prior probabilities of four causes of the explosion are 0.20, 0.40, 0.25 and 0.15. Based on all this information, what is

   a) the most likely cause of the explosion;
   b) the least likely cause of the explosion?

Solution:

St: static electricity
Ma: malfunctioning of equipment
Ca: carelessness
Sa: sabotage

$$\left.\begin{array}{l} P\left(E\middle|St\right)=0.25, \quad P\left(E\middle|Ma\right)=0.20, \\ P\left(E\middle|Ca\right)=0.40, \quad P\left(E\middle|Sa\right)=0.75 \end{array}\right. \text{likelihoods}$$

$P(St)$=0.20, $P(Ma)$=0.40, $P(Ca)$=0.25, $P(Sa)$=0.15 priors

$$\begin{aligned} P(E) &= P\left(E\middle|St\right)P(St)+P\left(E\middle|Ma\right)P(Ma)+P\left(E\middle|Ca\right)P(Ca)+P\left(E\middle|Sa\right)P(Sa) \\ &= (0.25)(0.20)+(0.20)(0.40)+(0.40)(0.25)+(0.75)(0.15) \\ &= 0.3425 \end{aligned}$$

$$P\left(St\middle|E\right)=\frac{P\left(E\middle|St\right)P(St)}{P(\mathrm{E})} = \frac{(0.25)(0.20)}{0.3425} \cong 0.146 \qquad P\left(Ma\middle|E\right)=\frac{P\left(E\middle|Ma\right)P(Ma)}{P(\mathrm{E})} = \frac{(0.20)(0.40)}{0.3425} \cong 0.2336$$

$$P\left(Ca\middle|E\right)=\frac{P\left(E\middle|Ca\right)P(Ca)}{P(\mathrm{E})} = \frac{(0.25)(0.40)}{0.3425} \cong 0.292 \qquad P\left(Sa\middle|E\right)=\frac{P\left(E\middle|Sa\right)P(Sa)}{P(E)} = \frac{(0.75)(0.15)}{0.3425} = 0.3285$$

a) the most likely cause of the explosion at a construction site occurred as the result of **sabotage** with probability of 0.3285.
b) the least likely cause of the explosion at a construction site occurred as the result of **static electricity** with probability of 0.146.

4) A polygraph (lie detector) is said to be 90% reliable in the following sense: There is a 90% chance that a person who is telling the truth will pass the polygraph test; and there is a 90% chance that a person telling a lie will fail the polygraph test.

a) Suppose a population consists of 5% liars. A random person takes a polygraph test, which concludes that he/she is lying. What is the probability that he/she is actually lying?
b) Consider the probability that a person is actually lying given that the polygraph says that he/she is. Using the definition of reliability, how reliable must the polygraph test be in order that this probability is at least 85%?

Solution:

Po={ A polygraph says the person is liar} L={the person is actually liar}

a)

$$P\left(Po\middle|L\right)=0.90, \quad P\left(\bar{Po}\middle|\bar{L}\right)=0.90, \quad P(L)=0.05, \quad P(\bar{L})=0.95$$

$$P\left(L \mid Po\right)=\frac{P\left(Po \mid L\right)P(L)}{P\left(Po \mid L\right)P(L)+P\left(Po \mid \bar{L}\right)P(\bar{L})}$$

$$=\frac{(0.90)(0.05)}{(0.90)(0.05)+(0.10)(0.95)}=\frac{0.045}{0.14}=0.3214$$

b)

$$P\left(L \mid Po\right)=\frac{P\left(Po \mid L\right)P(L)}{P\left(Po \mid L\right)P(L)+P\left(Po \mid \bar{L}\right)P(\bar{L})}$$

$$0.85=\frac{(p)(0.05)}{(p)(0.05)+(1-p)(0.95)}$$

$$0.8075=0.815p$$

$$p=\frac{0.8075}{0.815}=0.9908$$

If the probability that a person is actually lying given that the polygraph says that he/she is increases, the reliability will increase. It means that there is a 99.08% chance that a person who is telling the truth will pass the polygraph test; and there is a 99.08% chance that a person telling a lie will fail the polygraph test.

5) Your friend has three dice. One die is fair. One die has fives on all six sides. One die has fives on three sides and four on three sides. A die is chosen at random. It comes up five. Find the probability that the chosen die is the fair one.

Solution:
One die is fair, F={fair die}
All sides are five, A={all sides of die are five}
Three sides are five, three sides are four T={ three sides of are five, three sides of die are four}
Die comes up five B={the face up of the die is five}

$$P\left(F \mid B\right)=\frac{P\left(B \mid F\right)P(F)}{P\left(B \mid F\right)P(F)+P\left(B \mid A\right)P(A)+P\left(B \mid T\right)P(T)}$$

$$=\frac{(1/6)(1/3)}{(1/6)(1/3)+(1)(1/3)+(1/2)(1/3)}=\frac{1}{10}$$