Actuary Probability I

Probability Theory II

Irving Gómez Méndez

August-December, 2021

Conditional Probability

Conditional Probability, Introduction

Consider a population of 20 students, 14 study Medicine and 6 study Engineering. We select at random without replacing two students and consider the events:

E_1 : “The first student studies Medicine.”

E_2 : “The second student studies Engineering.”

The sample space consists of the collection of order pairs $(a_i, a_j); (a_i, b_k); (b_k, a_i); (b_k, b_h)$ where a_i are students of Medicine and b_j are students of Engineering, $i \neq j; k \neq h; i, j \leq 14; h, k \leq 6$.

The number of elementary events is 20×19. The next table shows the number of sample points that correspond to the partition of Ω accordingly yo the events E_1, E_2 and their complements. The last column and row show the total.

	E_2^c	E_2	
E_1	14×13	14×6	14×19
E_1^c	6×14	6×5	6×19
	14×19	6×19	20×19

Using this table, it is easy to calculate probabilities like

$$\mathbb{P}(E_1 \cap E_2) = \frac{14 \times 6}{20 \times 19};$$

$$\mathbb{P}(E_1) = \frac{14 \times 19}{20 \times 19}; \quad \mathbb{P}(E_1^c \cap E_2) = \frac{6 \times 5}{20 \times 19}.$$

Consider the next problem: If we know that the first student studies Medicine, what is the probability that second student studies Medicine too?

In this case, we can see from the table that there are 14×19 possible results, from which 14×13 are favorable to the event E_2, thus the probability of the event of interest is

$$\frac{14 \times 13}{14 \times 19} = \frac{(14 \times 13)/(20 \times 19)}{(14 \times 19)/(20 \times 19)} = \frac{\mathbb{P}(E_1 \cap E_2^c)}{\mathbb{P}(E_1)}$$

The probability that we calculated is called "*conditional probability of E_2 given E_1*" and is denoted as $\mathbb{P}(E_2|E_1)$

Observe that $\mathbb{P}(E_2^c) = \frac{14\times 19}{20\times 19} = \frac{7}{10}$ does not coincide $\mathbb{P}(E_2^c|E_1) = \frac{13}{19}$. Since we know that E_1 happened we have additional information that modifies the sample space: the new population, for the second selection does not coincide with the original, since there are 13 students of medicine form a total of 19 possible students.

Note that if the sampling is done with replacement, the result of the first extraction does not gives any information for the second. In this case:

$$\mathbb{P}(E_2^c|E_1) = \mathbb{P}(E_2^c) = \frac{7}{10}$$

Conditional Probability

Let be $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space and let be $B \in \mathcal{A}$ such that $\mathbb{P}(B) > 0$. Define a new function $\mathbb{P}(\cdot|B : \mathcal{A} \to \mathbb{R})$ as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{for all } A \in \mathcal{A}.$$

This function is a (measure of) probability; since:

1. $\mathbb{P}(A|B) \geq 0$ for all $A \in \mathcal{A}$.
2. $\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$.
3. Let be $A_1, A_2, \ldots$ disjoint sets in $\mathcal{A}$, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n | B\right) = \frac{\mathbb{P}\left(B \cap \bigcup_{n=1}^{\infty} A_n\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup_{n=1}^{\infty} B \cap A_n\right)}{\mathbb{P}(B)}$$

$$= \frac{\sum_{n=1}^{\infty} \mathbb{P}(B \cap A_n)}{\mathbb{P}(B)} = \sum_{n=1}^{\infty} \mathbb{P}(A_n|B).$$

Examples of Conditional Probability

1. We throw a dice twice.
   (a) If the sum of the results is 8, what is the probability that the first number was k, $1 \leq k \leq 6$?
   (b) If the first number was 3, what is the probability that the second was k, $1 \leq k \leq 6$?
   (c) If the first number was 3, what is the probability that the sum of both is 7?

   Let be X the result of the first throw and Y the result of the second. We know that $\mathbb{P}(X = k) = \mathbb{P}(Y = k) = 1/6$, $1 \leq k \leq 6$.

(a). We want to calculate

$$\mathbb{P}(X = k|X + Y = 8) = \frac{\mathbb{P}((X = k) \cap (X + Y = 8))}{\mathbb{P}(X + Y = 8)}$$

For the probability in the denominator, observe that there are 5 results whose sum is 8 from a total of 36 possible results, which correspond to the order pairs $(2, 6); (3, 5); (4, 4); (5, 3); (6, 2)$, so the probability in the denominator is $5/36$. On the other hand, the probability in the numerator is zero if $k = 1$. For $2 \leq k \leq 6$ there is just one result for the second throw which is $Y = 8 - k$, so the probability in the numerator is $1/36$. Finally, we have

$$\mathbb{P}(X = k|X + Y = 8) = \begin{cases} 1/5 & \text{if } 2 \leq k \leq 6, \\ 0 & \text{if } k = 1. \end{cases}$$

(b). We want to calculate

$$\mathbb{P}(Y = k|X = 3) = \frac{\mathbb{P}((Y = k) \cap (X = 3))}{\mathbb{P}(X = 3)}$$

We know that $\mathbb{P}(X = 3) = 1/6$. To find the probability in the numerator observe that from 36 possible results just one correspond to the event $(Y = k) \cap (X = 3)$, thus that probability is $1/36$. Hence

$$\mathbb{P}(Y = k|X = 3) = \frac{1/36}{1/6} = \frac{1}{6}$$

This result is equal to $\mathbb{P}(Y = k)$ which coincides with the intuition, since knowing the result of the first throw does not affect in any way the second.

(c). In this case, we are interested in

$$\mathbb{P}(X+Y=7|X=3) = \frac{\mathbb{P}(X+Y=7)\cap(X=3)}{\mathbb{P}(X=3)},$$

but

$$(X+Y=7)\cap(X=3) = (3+Y=7)\cap(X=3) = (Y=4)\cap(X=3),$$

so

$$\begin{aligned}\mathbb{P}(X+Y=7|X=3) &= \frac{\mathbb{P}((X+Y=7)\cap(X=3))}{\mathbb{P}(X=3)}\\ &= \frac{\mathbb{P}((Y=4)\cap(X=3))}{\mathbb{P}(X=3)}\end{aligned}$$

and from the result of (b) we know that this probability is 1/6.

2. We throw two dice until the sum is 7 or 8. If the sum is 7 player A wins, if it is 8 player B wins. What is the probability that A wins?

   To solve this problem, note that the probability that A wins is the conditional probability that the sum is 7, given that the sum has been 7 or 8, that is

$$\mathbb{P}(A) = \frac{\mathbb{P}(\{7\})}{\mathbb{P}(\{7,8\})} = \frac{\frac{6}{36}}{\frac{11}{36}} = \frac{6}{11}.$$

3. We consider now a frequent situation in quality control and medical prevention.

   To control some disease in a population where the proportion of sick people is p it is used a medical exam to detect the illness. It is known that the probability that the exam detect the illness from a sick person is 0.90, and the probability that the exam classify a healthy person as sick is 0.01. Find the probability that a person is sick when the medical exam says so.

To answer the question, we choose at random a person and consider the events:
S :"the person is sick." R :"the exam detects the person as sick."
We want to calculate

$$\mathbb{P}(S|R) = \frac{\mathbb{P}(S \cap R)}{\mathbb{P}(R)}$$

we know that

$$\mathbb{P}(S) = p,$$
$$\mathbb{P}(R|S) = \frac{\mathbb{P}(S \cap R)}{\mathbb{P}(S)} = 0.90, \mathbb{P}(R|S^c) \quad = \frac{\mathbb{P}(S^c \cap R)}{\mathbb{P}(S^c)} = 0.01.$$

From the first two equations, we have

$$\mathbb{P}(S \cap R) = 0.90p$$

and from the first and third

$$\mathbb{P}(S^c \cap R) = 0.01(1 - p)$$

Hence,

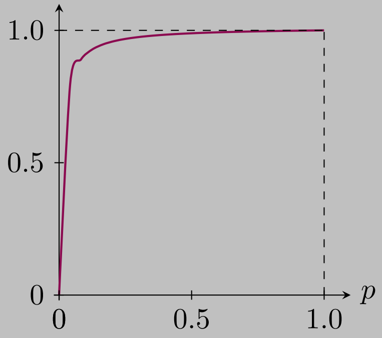
$$\mathbb{P}(R) = \mathbb{P}(S \cap R) + \mathbb{P}(S^c \cap R) \qquad = 0.90p + 0.01(1-p)$$

and

$$\mathbb{P}(S|R) = \frac{0.90p}{0.90p + 0.01(1-p)} = \frac{90p}{89p+1}$$

In particular if $p = 1/30$, $P(S|R) \approx 0.76$.

Considering $P(S|R)$ as function of p we observe that if the proportion p of sick people is small, the method of massive control is insufficient, since $P(S|R)$ is far of 1. For example, if $p = 0.001$, $P(S|R) \approx 0.083$



Properties of Conditional Probability

Properties of Conditional Probability

1. If A and B are disjoint, then $\mathbb{P}(A|B) = 0$.
   Indeed,

$$A \cap B = \varnothing \Rightarrow \mathbb{P}(A \cap B) = 0 \quad \text{and} \quad \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

2. If $B \subset A$, then $\mathbb{P}(A|B) = 1$, since $\mathbb{P}(A \cap B) = \mathbb{P}(B)$.
3. **Multiplication Law**. For any finite collection of events $A_1, \ldots, A_n$, we have

$$\begin{aligned}&\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n)\\ =\ &\mathbb{P}(A_1|A_2 \cap A_3 \cap \cdots \cap A_n)\mathbb{P}(A_2|A_3 \cap \cdots \cap A_n) \cdots \mathbb{P}(A_{n-1}|A_n)\mathbb{P}(A_n)\end{aligned}$$

   Always that $\mathbb{P}(A_2 \cap A_3 \cap \cdots \cap A_n) > 0$.

Proof
Because

$$\mathbb{P}(A_n) \geq \mathbb{P}(A_{n-1} \cap A_n) \geq \cdots \geq \mathbb{P}(A_2 \cap A_3 \cap \cdots \cap A_n) > 0$$

all the conditional probabilities are well defined. If we explicitly write the right-hand side of the equation, we have

$$\frac{\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n)}{\mathbb{P}(A_2 \cap \cdots \cap A_n)} \frac{\mathbb{P}(A_2 \cap \cdots \cap A_n)}{\mathbb{P}(A_3 \cap \cdots \cap A_n)} \cdots \frac{\mathbb{P}(A_{n-1} \cap A_n)}{\mathbb{P}(A_n)} \mathbb{P}(A_n)$$

and simplifying we have the left.hand side of the equation.

Examples of Conditional Probability

4. We select at random without replacement three cards from a pack of 52 cards. Find the probability that none of the cards is spade.

   Let be A_i: "The i-th card is not spade". We want to calculate

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2 \cap A_3) &= \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_2 \cap A_1)\\ &= \frac{39}{52}\frac{38}{51}\frac{37}{50} \approx 0.4135\end{aligned}$$

Law of Total Probability

Law of Total Probability

Le be $B_1, B_2, \ldots$ a finite or numerable family of pairwise disjoint sets whose union is Ω. Then, for any event A,

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

where the sum is over all the indexes i such that $\mathbb{P}(B_i) > 0$.

Proof

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) = \mathbb{P}(A \cap (\cup_i B_i)) \\ &= \mathbb{P}(\cup_i (A \cap B_i) = \sum_i \mathbb{P}(A \cap B_i) \\ &= \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i).\end{aligned}$$

Examples of Conditional Probability

5. There are n cards with names in a bag, and we select two, successively and without replacement. If $m < n$ of the cards have female names, find the probability that the second card has a female name.

   Let be A the event whose probability we want to find and F:"The first card has a female name". The events F and F^c are disjoint and form a partition of the sample space.

Thus,

$$\mathbb{P}(A) = \mathbb{P}(A|F)\mathbb{P}(F) + \mathbb{P}(A|F^c)\mathbb{P}(F^c)$$

and

$$\mathbb{P}(F) = \frac{m}{n}; \mathbb{P}(F^c) = \frac{n-m}{n}; \mathbb{P}(A|F) = \frac{m-1}{n-1}; \mathbb{P}(A|F^c) = \frac{m}{n-1}$$

Therefore,

$$\begin{aligned}\mathbb{P}(A) &= \frac{(m-1)}{(n-1)}\frac{m}{n} + \frac{m}{(n-1)}\frac{(n-m)}{n} \\ &= \frac{m}{(n-1)n}(m-1+n-m) = \frac{m}{n}\end{aligned}$$

6. What is the probability of getting 6 distinct numbers after throwing 6 dice?

Consider the events:

E_1: "The first dice shows any number."

E_2: "The number in the second dice is different from the first."

E_3: "The number in the third dice is different from the first and second."

And so on.

Then, we have

$$\mathbb{P}(E_1) = 1$$
$$\mathbb{P}(E_2|E_1) = 5/6$$
$$\vdots$$
$$\mathbb{P}(E_6|E_1 \cap E_2 \cap \cdots \cap E_5) = 1/6$$

therefore,

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_6) = 1 \times 5/6 \times \cdots 1/6 = \frac{6!}{6^6} \approx 0.015$$

7. We select two cards from a pack of 52 cards. Find the probability that the selected cards are an ace and a 10.

Consider the events:
A_1: "The first card is an ace".
A_2: "The second card is an ace". B_1: "The first card is a 10".
B_2: "The second card is a 10".
C: "We select an ace and a 10".
Thus

$$\begin{aligned}\mathbb{P}(C) &= \mathbb{P}(B_2|A_1)\mathbb{P}(A_1) + \mathbb{P}(A_2|B_1)\mathbb{P}(B_1)\\ &= \frac{4}{51}\frac{4}{52} + \frac{4}{51}\frac{4}{52}\\ &= \frac{8}{663}\end{aligned}$$

8. **Craps**. The game of *craps* has the next rules. The player throws two dice, if the result is 7 or 11, he/she wins. If it is 2, 3 or 12, he/she looses. If the sum is any other number, that number becomes his/her target and from that moment the player throws the dice until they sum his/her target, in such case the player wins, or when the sum is 7, in such case the player looses. What is the probability of winning this game?

Let be A_j the event of getting j in the first throw, with $j = 2, 3, \ldots, 12$ and G the event that the player wins. By law of total probability we have

$$\mathbb{P}(G) = \sum_{j=2}^{12} \mathbb{P}(G|A_j)\mathbb{P}(A_j)$$

We know that if $j = 2, 3$ or 12, the player looses, so the corresponding terms in the sum are 0. If $j = 7$ or 11, the player wins, so $\mathbb{P}(G|A_j) = 1$ for $j = 1, 7$. For the rest of the results we need to calculate the conditional probability $\mathbb{P}(G|A_j)$ for $j = 4, 5, 6, 8, 9, 10$.
From a previous example, we know that

$$\mathbb{P}(G|A_j) = \mathbb{P}(\{j\}|\{j, 7\}) = \frac{\mathbb{P}(\{j\})}{\mathbb{P}(\{j, 7\})}$$

Using the last formula, we have

$$\mathbb{P}(G|A_4) = \mathbb{P}(G|A_{10}) = \frac{1}{3}; \quad \mathbb{P}(G|A_5) = \mathbb{P}(G|A_9) = \frac{2}{5};$$

$$\mathbb{P}(G|A_6) = \mathbb{P}(G|A_8) = \frac{5}{11}$$

Therefore,

$$\begin{aligned}
\mathbb{P}(G) &= \sum_{j=4}^{11} \mathbb{P}(G|A_j)\mathbb{P}(A_j) \\
&= \frac{1}{3} \times \frac{3}{36} + \frac{2}{5} \times \frac{4}{36} + \frac{5}{11} \times \frac{5}{36} + 1 \times \frac{6}{36} \\
&\quad + \frac{5}{11} \times \frac{5}{36} + \frac{2}{5} \times \frac{4}{36} + \frac{1}{3} \times \frac{3}{36} + 1 \times \frac{2}{36} \approx 0.4929
\end{aligned}$$

Bayes' Theorem

Bayes' Theorem

Let be A, B events and $\mathbb{P}(B) \neq 0$, then

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

Furthermore, let be $B_1, B_2, \ldots$ a finite or numerable partition of Ω and let be A any other event such that $\mathbb{P}(A) > 0$. Then

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)}{\sum_j \mathbb{P}(A|B_j)\mathbb{P}(B_j)}$$

Examples of Bayes' Theorem

1. From 100 patients in a hospital with certain disease,10 are chosen for a treatment that augments the probability of heal from 0.5 to 0.75. Time after, a medic finds a healed patient, what is the probability that the patient received the treatment?

   Let be
   H : "the patient is healed".
   T : "the patient received the treatment".

From the information given we have

$$\mathbb{P}(T) = \frac{10}{100} = 0.1; \quad \mathbb{P}(T^c) = \frac{90}{100} = 0.9$$

$$\mathbb{P}(H|T) = 0.75; \quad \mathbb{P}(H|T^c) = 0.5$$

Using the Bayes' theorem

$$\mathbb{P}(T|H) = \frac{\mathbb{P}(H|T)\mathbb{P}(T)}{\mathbb{P}(H|T)\mathbb{P}(T) + \mathbb{P}(H|T^c)\mathbb{P}(T^c)}$$
$$= \frac{0.75 \times 0.1}{0.75 \times 0.1 + 0.5 \times 0.9} = \frac{1}{7}$$

2. Three boxes contain two coins each one. In the first, B_1, both are coins of gold; in the second, B_2, both are coins of silver and in the third, B_3, one coin is of gold and one coin is of silver. We pick one box at random and then one coin also at random. If the coin is of gold, what is the probability that it comes from the box with two gold coins?

   We know that $\mathbb{P}(B_i) = \frac{1}{3}$. Let be G : "we select a gold coin". Using Bayes' theorem

$$\begin{aligned}\mathbb{P}(B_1|G) &= \frac{\mathbb{P}(G|B_1)\mathbb{P}(B_1)}{\mathbb{P}(G|B_1)\mathbb{P}(B_1) + \mathbb{P}(G|B_2)\mathbb{P}(B_2) + \mathbb{P}(G|B_3)\mathbb{P}(B_3)} \\ &= \frac{1 \times \frac{1}{3}}{1 \times \frac{1}{3} + 0 \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3}} = \frac{2}{3}\end{aligned}$$

3. Three mutually exclusive diseases A, B and C have the same symptoms H. Accordingly to a clinical study the probabilities of getting the diseases are 0.01; 0.005 and 0.02, respectively. Furthermore, the probability that a patient shows the symptoms H for each disease are 0.90; 0.95 and 0.75, respectively. If a sick person has the symptoms H, what is the probability that has the disease A?

We know that

$$\mathbb{P}(A) = 0.01; \quad \mathbb{P}(B) = 0.005; \quad \mathbb{P}(C) = 0.02$$

$$\mathbb{P}(H|A) = 0.90; \quad \mathbb{P}(H|B) = 0.95 \quad \mathbb{P}(H|C) = 0.75$$

using Bayes' theorem

$$\mathbb{P}(A|H) = \frac{\mathbb{P}(H|A)\mathbb{P}(A)}{\mathbb{P}(H|A)\mathbb{P}(A) + \mathbb{P}(H|B)\mathbb{P}(B) + \mathbb{P}(H|C)\mathbb{P}(C)}$$

$$= \frac{0.90 \times 0.01}{0.90 \times 0.01 + 0.95 \times 0.005 + 0.75 \times 0.02} = \frac{2}{3} \approx 0.313$$

4. A student answers a multiple-select question with four possible options. Assume that the probability that the student knows the answer to the question is 0.8 and the probability that he/she guesses is 0.2. If the student guess, the probability of selecting the correct answer is 0.25. If the student answered correctly, what is the probability that the student really knew the answer?

Define the next events:
K : "the student knows the answer".
C : "the student answers correctly".
We know that

$$\mathbb{P}(K) = 0.8; \quad \mathbb{P}(C|K) = 1; \quad \mathbb{P}(C|K^c) = 0.25$$

using Bayes' theorem

$$\begin{aligned}\mathbb{P}(K|C) &= \frac{\mathbb{P}(C|K)\mathbb{P}(K)}{\mathbb{P}(C|K)\mathbb{P}(K) + \mathbb{P}(C|K^c)\mathbb{P}(K^c)} \\ &= \frac{1 \times 0.8}{1 \times 0.8 + 0.25 \times 0.2} = 0.941\end{aligned}$$

5. **The Monty Hall problem**. Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say door 1, and the host, who knows what's behind the doors, opens another door, say door 3, which has a goat. He says to you, 'Do you want to pick door 2?' Is it to your advantage to switch your choice of doors?

Define the next events:
H : “the host opens door 3”.
D_i : “the car is behind door i”.
Given that the player have chosen door 1, from the rules of the game, we know that

$$\mathbb{P}(D_i) = \frac{1}{3}; \quad \mathbb{P}(H|D_1) = \frac{1}{2}; \quad \mathbb{P}(H|D_2) = 1; \quad \mathbb{P}(H|D_3) = 0.$$

We want to calculate $\mathbb{P}(D_1|H)$ and $\mathbb{P}(D_2|H)$. Using Bayes' theorem to find the first probability, we have

$$\begin{aligned}\mathbb{P}(D_1|H) &= \frac{\mathbb{P}(H|D_1)\mathbb{P}(D_1)}{\mathbb{P}(H|D_1)\mathbb{P}(D_1) + \mathbb{P}(H|D_2)\mathbb{P}(D_2) + \mathbb{P}(H|D_3)\mathbb{P}(D_3)} \\ &= \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{1}{3}\end{aligned}$$

Analogously,

$$\begin{aligned}\mathbb{P}(D_2|H) &= \frac{\mathbb{P}(H|D_2)\mathbb{P}(D_2)}{\mathbb{P}(H|D_1)\mathbb{P}(D_1) + \mathbb{P}(H|D_2)\mathbb{P}(D_2) + \mathbb{P}(H|D_3)\mathbb{P}(D_3)} \\ &= \frac{1 \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{2}{3}\end{aligned}$$

Therefore, by switching, the player double their chances from $\frac{1}{3}$ to $\frac{2}{3}$.

One as the player might think in the next way: After I have selected a door, one of the remaining doors is going to have a goat. And I already know that the host is going to open a door with a goat behind it. Thus, when the host opens a door with a goat behind it, he/she is giving me no extra information. Hence, the probability that 'the car is behind door 1' or 'the car is behind door 2' have the same probability, i.e. 1/2 each one.

We are now facing a paradox, what is wrong with this way of reasoning? and what has created this apparent contradiction?

6. Suppose you take a medical test to see if you have a disease, and it comes back positive. How likely is it that you have the disease? For specificity, let's say the disease is breast cancer, and the test is a mammogram.

   In this example the *forward* probability is the probability of a positive test, given that you have the disease: $\mathbb{P}(\text{test}|\text{disease})$. This is what a doctor would call the "sensitivity" of the test, or its ability to correctly detect an illness. Generally it is the same for all types of patients, because it depends only on the technical capability of the testing instrument to detect the abnormalities associated with the disease.

The *inverse* probability is the one you surely care more about: What is the probability that I have the disease, given that the test came out positive? This is $\mathbb{P}(\text{disease}|\text{test})$. This probability is not necessarily the same for all types of patients; we would certainly view the positive test with more alarm in a patient with a family history of the disease than in one with no such history.

Suppose a forty-year-old woman gets a mammogram to check for breast cancer, and it comes back positive. The hypothesis, D (for "disease"), is that she has cancer. The evidence, T (for "test"), is the result of the mammogram. How strongly should she believe the hypothesis?

We can answer this question by rewriting Baye's rule as follows:

$$\mathbb{P}(D|T) = (\text{likelihood tratio}) \times (\text{prior probability})$$

where the new term "likelihood ratio" is given by $\mathbb{P}(T|D)/\mathbb{P}(T)$. It measures how much more likely the positive test is in people with the disease than the general population. This formula tells us that the new evidence T augments the probability of D by a fixed ratio, no matter what the prior probability was.
For a typical forty-year-old woman, the probability of getting breast cancer in the next year is about one in seven hundred, so we'll use that as our prior probability, that is $\mathbb{P}(D) = 1/700$ and $\mathbb{P}(D^c) = 699/700$.

To compute the likelihood ratio, we need to know $\mathbb{P}(T|D)$ and $\mathbb{P}(T)$. In the medical context, $\mathbb{P}(T|D)$ is the sensitivity of the mammogram–the probability that it will come back positive if you have cancer. According to the Breast Cancer Surveillance Consortium (BCSC), the sensitivity of mammograms for forty-year-old women is 73 percent, that is $\mathbb{P}(T|D) = 0.73$. The denominator $\mathbb{P}(T)$ is a but trickier. A positive test, T, can come both from patients who have the disease and from patients who don't. It can be computed through the formula

$$\mathbb{P}(T) = \mathbb{P}(T|D)\mathbb{P}(D) + \mathbb{P}(T|D^c)\mathbb{P}(D^c).$$

$\mathbb{P}(T|D^c)$ is known as the false positive rate. According to the BCSC, the false positive rate for forty-year-old women is about 12 percent, that is $\mathbb{P}(T|D^c) = 0.12$.

Hence,

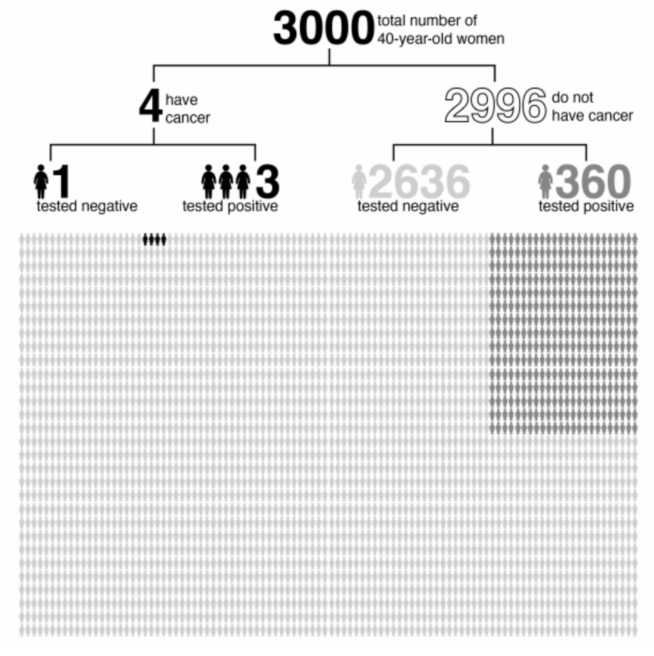
$$\mathbb{P}(T) = \left(\frac{1}{700}\right) \times 0.73 + \left(\frac{699}{700}\right) \times 0.12 \approx 0.121.$$

The likelihood ratio is

$$\frac{0.73}{0.121} \approx 6.$$

Since her prior probability was one |/700, her updated probability is

$$\mathbb{P}(D|T) = 6 \times \frac{1}{700} \approx 1/116.$$


3000 total number of 40-year-old women
4 have cancer
2996 do not have cancer
1 tested negative
3 tested positive
2636 tested negative
360 tested positive

However, the story would be very different if the patient has propensity to have cancer. Denote by H the event "the patient has propensity to develop cancer" then,

$$\begin{aligned}\mathbb{P}(D|T,H) &= \frac{\mathbb{P}(T|D,H)}{\mathbb{P}(T)}\mathbb{P}(D|H) \\ &= \frac{\mathbb{P}(T|D)}{\mathbb{P}(T)}\mathbb{P}(D|H) \\ &= 6\mathbb{P}(D|H)\end{aligned}$$

Say, for example that a woman has a gene that put her at high risk for breast cancer–say, $\mathbb{P}(D|H) = 1/18$. Then a positive test would increase the probability to $\mathbb{P}(D|T,H) = 1/3$.

Brief History of Bayes' Rule

Thomas Bayes was concerned with the probabilities of two events, one (the hypothesis) occurring before the other (the evidence). To set the context, in 1748, the Scottish philosopher David Hume had written an essay titled "On Miracles," in which he argued that eyewitness testimony could never prove that a miracle had happened. Hume's main point was that inherently fallible evidence cannot overrule a proposition with the force of natural law, such as "Dead people stay dead." For Bayes, this assertion provoked a natural question: How much evidence would it take to convince us that something we consider improbable has actually happened? When does a hypothesis cross the line from impossibility to improbability and even to probability or virtual certainty?

Bayes' rule acts as a normative rule for updating beliefs in response to evidence. It implies that the more surprising the evidence e–that is, the smaller $\mathbb{P}(e)$ is–the more convinced one should become of its case H.
For example if e is a miracle ("Christ rose from the dead"), and H is a closely related hypothesis ("Christ is the son of God"), our degree of belief in H is very dramatically increased if we know for a fact that e is true. The more miraculous the miracle, the more credible the hypothesis that explains its occurrence.

$$\mathbb{P}(H|e) = \frac{\mathbb{P}(H \cap e)}{\mathbb{P}(e)} = \frac{\overbrace{\mathbb{P}(e|H)}^{\simeq 1}}{\mathbb{P}(e)}\mathbb{P}(H)$$

Independence

Independence

We say that two events A and B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

If A and B are independent then,

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

That is, the occurrence of A does not give information on the occurrence of B.

If A and B are independent, so do a) A and B^c; b) A^c and B^c

$$\begin{aligned}\mathbb{P}(A^c \cap B^c) &= \mathbb{P}((A \cup B)^c) = 1 - \mathbb{P}(A \cup B) \\ &= 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A \cap B) \\ &= [1 - \mathbb{P}(A)][1 - \mathbb{P}(B)] \\ &= \mathbb{P}(A^c)\mathbb{P}(B^c).\end{aligned}$$

Examples of Independence

1. A stock of 10 objects has 4 defective products and 6 of good quality. Two objects are extracted successively without replacement. Let be the events D_1 : "the first object is defective" and D_2 : "the second object is defective". Are these events independent? What happens if the objects are extracted with replacement?

   In the first case we can find $\mathbb{P}(D_2)$ as:

   $$\begin{aligned}\mathbb{P}(D_2) &= \mathbb{P}(D_2|D_1)\mathbb{P}(D_1) + \mathbb{P}(D_2|D_1^c)\mathbb{P}(D_1^c)\\ &= \frac{3}{9}\frac{4}{10} + \frac{4}{9}\frac{6}{10} = \frac{2}{5}\end{aligned}$$

   On the other hand,

   $$\mathbb{P}(D_2|D_1) = \frac{3}{9} \neq \frac{2}{5} = \mathbb{P}(D_2)$$

   so D_1 and D_2 are not independent.

If, the objects are selected with replacement, then $\mathbb{P}(D_1) = \mathbb{P}(D_2) = \frac{4}{10}$ and $\mathbb{P}(D_1 \cap D_2) = \left(\frac{4}{10}\right)^2$, so the events are independent. ▲

We can generalize the definition of independence to any family of events:
Let be $C = \{A_i,\, i \in I\}$ a family of events. We say that the events A_i are independent if for any finite collection of event $A_{i_1}, A_{i_2}, \ldots, A_{i_n} \in C$, it is satisfied:

$$\mathbb{P}\left(\bigcap_{j=1}^{n} A_{i_j}\right) = \prod_{j=1}^{n} \mathbb{P}\left(A_{i_j}\right)$$

In that case we say that C is a family of independent events.

Note that in the definition we only consider finite events of C, but we consider all the finite collections. For example, if the family has three events, it is not enough to verify the independent for every pair of events. Indeed, consider the experiment of throwing two dice and $C = \{A_1, A_2, A_3\}$, where
A_1 : "the first dice shows 6".
A_2 : "the second dice shows 1".
A_3 : "the sum of the dice is 7".
Clearly,

$$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \frac{1}{6}$$

and

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_2 \cap A_3) = \frac{1}{6} \times \frac{1}{6}$$

but

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{36} \neq \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6}$$

so the three events are not independent.

2. If A and B are independent events and the probability that both happen is 0.16, while the probability that none happen s 0.36, find $\mathbb{P}(A)$ and $\mathbb{P}(B)$.

We know that $\mathbb{P}(A \cap B) = 0.16$ and $\mathbb{P}((A \cup B)^c) = 0.36$, so

$$\begin{aligned}\mathbb{P}(A \cup B) &= 0.64 \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - 0.16.\end{aligned}$$

Thus, we have the next equations

$$\begin{aligned}\mathbb{P}(A) + \mathbb{P}(B) &= 0.8, \\ \mathbb{P}(A)\mathbb{P}(B) &= 0.16.\end{aligned}$$

From there, we obtain $\mathbb{P}(A) = \mathbb{P}(B) = 0.4$.

3. What is the probability of getting three 6 when throwing 8 dice?

The problem is equivalent to find the probability of getting three times 6 throwing 8 times the same dice. Let be E_i the event "we get a 6 in the i-th throw". Let's find the the probability that the first three are 6 and the others are not, i.e.

$$\mathbb{P}(A_1 \cap E_2 \cap E_3 \cap E_4^c \cap E_5^c \cap E_6^c \cap E_7^c \cap E_8^c)$$

by independence

$$\left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^5.$$

This is the probability of any set of 8 throws with exactly three 6. Since there are $\binom{8}{3}$ of this type, the probability of interest is

$$\binom{8}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^5.$$

4. The first son of a woman has hemophilia. There are no previous cases of hemophilia in the woman's family, so the woman thinks that her son did not inherit the hemophilia from her and that the disease but it has been a mutation. Therefore, the probability that a second son has hemophilia should be a mutation again, which is a small number, m ($m = 10^{-5}$). What is the probability that a second son has hemophilia if the first one has hemophilia?

Define the events:
A: "the mother carrier the disease".
H_1: "the first son has hemophilia".
H_2: "the second son has hemophilia".

A man has chromosome XY and has hemophilia if and only if instead of the chromosome X has a chromosome X' with a gene that produces the hemophilia. Let be m the probability that a chromosome X mutates into a chromosome X'.

The mother has two chromosome X and the event A happens if at least one of those chromosomes mutates, which happens with probability

$$\mathbb{P}(A) = 1 - (1-m)^2 = 2m - m^2 \approx 2m$$

where we have assumed that mutations take place independently, and we have discarded the term m^2 since it is much smaller than $2m$.
If the mother carriers the disease and one of the chromosomes is X' her son will have probability $2/2$ of inherit the chromosome X, that is

$$\mathbb{P}(H_1|A) = \mathbb{P}(H_1^c|A) = \frac{1}{2}$$

On the other hand, if the mother does not carrier the disease, her son will have hemophilia if the chromosome X mutates:

$$\mathbb{P}(H_1|A^c) = m.$$

Moreover, by independence, we have

$$\mathbb{P}(H_2|A \cap H_1) = \mathbb{P}(H_2^c|A \cap H_1) = \mathbb{P}(H_2|A \cap H_1^c) = \mathbb{P}(H_2^c|A \cap H_1^c) = \frac{1}{2}$$

$$\mathbb{P}(H_2|A^c \cap H_1) = \mathbb{P}(H_2|A^c \cap H_1^c) = m.$$

We want to calculate

$$\mathbb{P}(H_2|H_1) = \frac{\mathbb{P}(H_1 \cap H_2)}{\mathbb{P}(H_1)}$$

$$\begin{aligned}\mathbb{P}(H_1 \cap H_2) &= \mathbb{P}(A \cap H_1 \cap H_2) + \mathbb{P}(A^c \cap H_1 \cap H_2) \\ &= \mathbb{P}(H_2|H_1 \cap A)\mathbb{P}(H_1|A)\mathbb{P}(A) + \mathbb{P}(H_2|H_1 \cap A^c)\mathbb{P}(H_1|A^c)\mathbb{P} \\ &\approx 2m\left(\frac{1}{2}\right)^2 + (1-2m)m^2 \\ &= \frac{m}{2} + m^2 - 2m^3 \approx \frac{m}{2},\end{aligned}$$

on the other hand

$$\begin{aligned}\mathbb{P}(H_1) &= \mathbb{P}(H_1|A)\mathbb{P}(A) + \mathbb{P}(H_1|A^c)\mathbb{P}(A^c) \\ &\approx 2m\left(\frac{1}{2}\right) + m(1-2m) \\ &\approx 2m.\end{aligned}$$

Therefore,

$$\mathbb{P}(H_2|H_1 \approx \frac{m/2}{2m}) = \frac{1}{4}.$$

5. **The problem of the points**. Independent trials resulting in a success with probability p and a failure with probability $1 - p$ are performed. What is the probability that n successes occur before m failures? If we think of A and B as playing a game such that A gains 1 point when a success occurs and B gains 1 point when a failure occurs, then the desired probability is the probability that A would win if the game were to be continued in a position where A needed n and B needed m more points to win.

This problem was posed to the French mathematician Blaise Pascal in 1654 by the Chevalier de Méré, who was a professional gambler at that time. In attacking the problem, Pascal introduced the important idea that the proportion of the prize deserved by the competitors should depend on their respective probabilities of winning if the game were to be continued at that point. Pascal initiated a correspondence with the famous French Pierre de Fermat, who had a great reputation as a mathematician. This celebrated correspondence, dated by some as the birth date of probability theory.

Fermat argued that, in order for n successes to occur before m failures, it is necessary and sufficient that there be at least n successes in the first $m+n-1$ trials. (Even if the game were to end before a total of $m+n-1$ trials were completed, we could still imagine that the necessary additional trials were performed.) This is true, for if there are at least n successes in the first $m+n-1$ trials, there could be at most $m-1$ failures in those $m+n-1$ trials; thus, n successes would occur before m failures.

Hence, since the probability of exactly k successes in $m+n-1$ trials is

$$\binom{m+n-1}{k} p^k (1-p)^{m+n-1},$$

it follows that the desired probability of n successes before m failures is

$$\sum_{k=n}^{m+n-1} \binom{m+n-1}{k} p^k (1-p)^{m+n-1}.$$