Example 1

From *The Week.*

> Suppose one out of every million people is a terrorist (if anything, an overestimate), and you've got a machine that can determine whether someone is a terrorist with 99.9 percent accuracy. You've used the machine on Mr. X, and it gives a positive result. What are the odds that Mr. X is a terrorist? Here's the answer: a 0.1 percent chance — which is to say, the 99.9 percent accurate test will give you the wrong answer 99.9 percent of the time.[1]

Let $X = 1$ be the event that Mr. X is a terrorist with probability one in a million.

$$
\begin{aligned}
P(X{=}1) &= 0.000001 \\
P(X{=}0) &= 0.999999
\end{aligned}
$$

Let $M = 1$ be the event that the machine gives a positive result. The accuracy of the machine is a conditional probability. Given the condition that Mr. X is a terrorist ($X = 1$), the probability that the machine gives a positive result ($M = 1$) is 99.9 percent.

$$P(M{=}1 \mid X{=}1) = 0.999$$

The trick is to swap M and X. In other words, determine the conditional probability of $X = 1$ given that $M = 1$. A conditional probability can be "turned around" using the following formula.

$$P(X{=}1 \mid M{=}\,1) = \frac{P(M{=}1 \mid X{=}1)P(X{=}1)}{P(M{=}1)}$$

The denominator $P(M{=}1)$ can be obtained from the law of total probability.

$$P(M{=}1) = P(M{=}1 \mid X{=}1)P(X{=}1) + P(M{=}1 \mid X{=}0)P(X{=}0)$$

[1] Cooper, Ryan. *The simple math problem that blows apart the NSA's surveillance justifications.* http://theweek.com/articles/547119/simple-math-problem-that-blows-apart-nsas-surveillance-justifications

The term $P(M{=}1 \mid X{=}0)$ is the probability of a false positive. Since the machine is 99.9 percent accurate, the probability the machine is wrong is 0.1 percent.

$$P(M{=}1 \mid X{=}0) = 0.001$$

Hence by the law of total probability

$$\begin{aligned} P(M{=}1) &= P(M{=}1 \mid X{=}1)P(X{=}1) + P(M{=}1 \mid X{=}0)P(X{=}0) \\ &= 0.999 \times 0.000001 + 0.001 \times 0.999999 \\ &= 0.001001 \end{aligned}$$

The probability that the machine is correct when it signals a positive result is

$$\begin{aligned} P(X{=}1 \mid M{=}\, 1) &= \frac{P(M{=}1 \mid X{=}1)P(X{=}1)}{P(M{=}1)} \\ &= \frac{0.999 \times 0.000001}{0.001001} \\ &= 0.000998 \end{aligned}$$

So the probability is indeed about 0.1 percent.

Example 2

The following hypothetical problem is from an MIT online course in Economics. MIT is planning to subject admitted freshman to "At Home Test Kit for Illicit Drugs" which has been shown to be fairly reliable, in the sense that 90% of those using drugs test positive, while only 10% of those not using drugs test positive. Assume that 20% of the population in your age group actually uses illicit drugs. Alas, you test positive. What is the probability that you are truly and fairly busted?

Let X be the event that the test is positive. Let U be the event that the student uses drugs. The following are given.

$$\begin{aligned} P(X \mid U) &= 0.9 \\ P(X \mid \bar{U}) &= 0.1 \\ P(U) &= 0.2 \end{aligned}$$

This is what is sought: The conditional probability that the student uses drugs given that the test result is positive.

$$P(U \mid X) = \frac{P(X \cap U)}{P(X)}$$

The value of $P(X \cap U)$ is obtained from the definition of conditional probability.

$$P(X \cap U) = P(X \mid U)P(U) = 0.9 \times 0.2 = 0.18$$

The value of $P(X)$ is obtained from total probability.

$$P(X) = P(X \mid U)P(U) + P(X \mid \bar{U})P(\bar{U}) = 0.9 \times 0.2 + 0.1 \times 0.8 = 0.26$$

Hence

$$P(U \mid X) = \frac{P(X \cap U)}{P(X)} = \frac{0.18}{0.26} = 0.6923$$

Example 3

Another MIT Problem.

Group	Sample Size Enrolled	Number Placed in Jobs	Percentage Placed in Jobs
Tax Credit Voucher	247	32	13.0
Direct Rebate Voucher	299	38	12.7
Control	262	54	20.6
Total	808	124	15.3

1. Construct separate t-tests comparing each of the two treatment groups with the control group. Did the treatments have statistically significant effects? *Yes.* If so, in what direction? *The wrong direction.*

```
> TaxCredit = numeric(247)
> TaxCredit[1:32] = 1
> DirectRebate = numeric(299)
> DirectRebate[1:38] = 1
> Control = numeric(262)
> Control[1:54] = 1
> t.test(TaxCredit,Control,var.equal=TRUE)
```

```
        Two Sample t-test

data:  TaxCredit and Control
t = -2.3111, df = 507, p-value = 0.02123
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.14162910 -0.01147533
sample estimates:
mean of x mean of y
0.1295547 0.2061069

> t.test(DirectRebate,Control,var.equal=TRUE)

        Two Sample t-test

data:  DirectRebate and Control
t = -2.5317, df = 559, p-value = 0.01163
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1403224 -0.0177107
sample estimates:
mean of x mean of y
0.1270903 0.2061069
```

2. Test whether the employment rates in the two treatment groups differ from each other. *No statistical difference.*

```
> t.test(TaxCredit,DirectRebate,var.equal=TRUE)

        Two Sample t-test

data:  TaxCredit and DirectRebate
t = 0.0856, df = 544, p-value = 0.9318
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.05410503  0.05903374
sample estimates:
mean of x mean of y
0.1295547 0.1270903
```

3. Construct 95% confidence intervals for the two treatment-control contrasts.

```
data:  TaxCredit and Control
95 percent confidence interval:
 -0.14162910 -0.01147533

data:  DirectRebate and Control
95 percent confidence interval:
 -0.1403224 -0.0177107
```

4. Footnote 11 reports a chi-square statistic for independence between groups (tax credit, rebate, and control) and job placement. Explain how this was constructed.