Lets do some statistics!

- P(x) – the probability that event x will occur
- P(x V y) = P(x)+P(y)
- Conditional probability P(x|y)=P(x&y)/P(y)
- Bayesian Theorem:

$$
\begin{aligned}
P(h|e) &= \frac{P(h)\cdot P(e|h)}{P(e)} \\
&= \frac{P(h)\cdot P(e|h)}{P(e|h)\cdot P(h) + P(e|\neg h)\cdot P(\neg h)}
\end{aligned}
$$

Example Question

So-called “false positives” result when a test falsely or incorrectly reports a positive result. For example, a medical test screening for HIV antibodies may return a positive result indicating that the patient carries the human immunodeficiency virus (HIV) when in fact he or she does not. Bayes’s theorem can be used to determine the probability that a positive result is in fact a false positive. If a disease is rare, then the majority of positive results may be false positives, even if the test is highly accurate. In order to see this, use Bayes’s theorem to do the following calculation.

Suppose an HIV test is such that if the patient is indeed infected, it returns a positive result in 99% of the cases and if the patient is not infected, the test returns a negative result in 95% of the cases. Assume further that the disease is so rare that only one in a population of 100,000 is infected; in other words, a randomly selected patient has a 0.00001 prior probability of carrying the virus. What is the probability that a positive test result is a false positive?

You may find it helpful to use the following abbreviations:

d : patient has the disease

p : test result is positive

Steps to get it right:

1. Determine the probability of the hypothesis **H**.
2. Determine the probability of observing evidence **e** independently of H.
3. Determine the probability of the evidence **e** given the hypothesis **H** –**P(e|H)**
4. Determine the probability of the evidence **e** given the hypothesis **H** is false –**P(e|~H)**
5. Plug into Bayesian theorem.
6. Just repeat with the new **P(H|e)** as **H** and $\mathbf{e_2}$

- P(d) = 0.00001 or 1/100,000
- P(p|d) = 99/100
- P(~p|~d) = 95/100
- P(p|~d) = 5/100
- P(~d|p) = ?

$$\begin{aligned} P(h|e) &= \frac{P(h) \cdot P(e|h)}{P(e)} \\ &= \frac{P(h) \cdot P(e|h)}{P(e|h) \cdot P(h) + P(e|\neg h) \cdot P(\neg h)} \end{aligned}$$

$$\frac{99{,}999/100{,}000 * 5/100}{99{,}999/100{,}000 * 5/100 + 1/100{,}000 * 99/100}$$

0.9998

Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding? (http://stattrek.com/Lesson1/Bayes.aspx)

P(rain) = 5/365

P(~rain) = 360/365

P(prediction | rain) = .9

P(prediction | ~rain) = 0.1

$$\frac{P(\text{pred} \mid \text{rain})*P(\text{rain})}{P(\text{pred} \mid \text{rain})P(\text{rain}) + P(\text{pred} \mid \sim\text{rain})P(\sim\text{rain})}$$

$$\frac{5/365 * 9/10}{5/365 * 9/10 + 360/365 * 1/10}$$

0.111

Suppose that Bob can decide to go to work by one of three modes of transportation, car, bus, or commuter train. Because of high traffic, if he decides to go by car, there is a 50% chance he will be late. If he goes by bus, which has special reserved lanes but is sometimes overcrowded, the probability of being late is only 20%. The commuter train is almost never late, with a probability of only 1%, but is more expensive than the bus.

Suppose that Bob is late one day, and his boss wishes to estimate the probability that he drove to work that day by car. His boss knows that he almost always takes the commuter train to work, never takes the bus, but sometimes, 10% of the time, takes the car. What is the bosses probability that Bob drove to work that day, given that he was late?

(http://www.medicine.mcgill.ca/epidemiology/joseph/courses/epib-607/bayesex.pdf)

$P(car) = 0.1$

$P(train) = 0.9$

$P(bus) = 0$

$P(late \mid car) = 0.5$

$P(late \mid train) = 0.01$

$P(late \mid bus) = 0.2$

$$\frac{P(late \mid car) * P(car)}{P(late \mid car)P(car) + P(late \mid bus)P(bus) + P(late \mid train)P(train)}$$

$$\frac{(0.5*0.1)}{(0.5*0.1+0.2*0+0.01*0.9)} =$$

0.8475

How does all this connect back to epistemology?