Bayes' Theorem Examples

Jeffrey Arnold

January 28, 2016

Testing for rare diseases

Suppose there is a disease that affects only 1 in 10000. Fortunately, there is an accurate test for this disease. It produces the correct result 99% of the the time. Let D be the event that the person has the disease, $\neg D$ that they don't have the disease, T that the test is positive, and $\neg T$ that the test is not positive.

$$\Pr(D) = 0.0001$$
$$\Pr(T|D) = 0.99$$
$$\Pr(T|\neg D) = 0.99$$

a) Given a positive test, what is the probability that a person has the disease? Solve this using the method of trees in the *OpenIntro* book, and with Bayes' Rule?
b) Given the result in the previous question, should everyone be tested for this disease?
c) Suppose a person takes two tests. What is the probability of having the disease given two positive tests?
d) What is the minimum number of positive tests and no negative tests that someone need to have in order to have a greater than 50% probability of having disease?
e) Suppose that the disease is genetic, and the probability of Bob having the disease is 0.25. What is the probability that Bob has the disease given a positive test?

Using Bayes' theorem the probability that a person has a disease given that they test positive is

$$
\begin{aligned}
\Pr(D|T) &= \frac{\Pr(T|D)\Pr(D)}{\Pr(T)} \\
&= \frac{\Pr(T|D)\Pr(D)}{\Pr(T|D)\Pr(D) + \Pr(T|\neg D)\Pr(\neg D)} \\
&= \frac{0.0001 \times 0.99}{0.0001 \times 0.99 + 0.9999 \times 0.01}
\end{aligned}
$$

```
pD <- 0.0001
pTD <- 0.99
pTnotD <- 0.01
(pTD * pD) / (pTD * pD + pTnotD * (1 - pD))
```

```
## [1] 0.009803922
```

So given a positive test, there is less than a 1% chance that a person has the disease. This is quite small, and suggests that widescale testing would be overwhelmed by false positives.

We can also do this using a tree. The package *openintro* provides the function `treeDiag` to draw these,

```
library("openintro")
treeDiag(main = c("Disease?", "Test Result"),
        p1 = c(0.0001, 0.9999),
        p2 = list(c(0.99, 0.01), c(0.01, 0.99)),
        out1 = c("Yes", "No"),
        out2 = c("Positive", "Negative"))
```

1

If there is a followup test, and it is also positive, the probability that the person has the disease rises to almost 50%.

$$\begin{aligned}
\Pr(D|T_1\&T_2) &= \frac{\Pr(T_1\&T_2|D)\Pr(D)}{\Pr(T_1\&T_2)} \\
&= \frac{\Pr(T_1|D)\Pr(T_2|D)\Pr(D)}{\Pr(T|D)\Pr(T_1|D)\Pr(T_2|D)\Pr(D) + \Pr(T_1|\neg D)\Pr(T_2|\neg D)\Pr(\neg D)} \\
&= \frac{0.0001 \times 0.99^2}{0.0001 \times 0.99^2 + 0.9999 \times 0.01^2}
\end{aligned}$$

```
(pTD ^ 2 * pD) / (pTD ^ 2 * pD + pTnotD^2 * (1 - pD))
```

```
## [1] 0.495
```

Note that this calculation depends on the tests being independent. An example of the tests not being independent is if the person has a genetic mutatation that is similar to but not the same as the one that causes the disease, but is flagged by the test.

Note that we get the same result by calculating the probabilities sequentially, using the results of the first

$$\begin{aligned}
\Pr(D|T_2) &= \frac{\Pr(T_2|D)\Pr(D|T_1)}{\Pr(T_2)} \\
&= \frac{\Pr(T_2|D)\Pr(D|T_1)}{\Pr(T_2|D)\Pr(D|T_1) + \Pr(T_2|\neg D)\Pr(\neg D|T_1)} \\
&\approx \frac{0.01 \times 0.99}{0.01 \times 0.99 + 0.99 \times 0.01}
\end{aligned}$$

```
pD_T1 <- (pTD * pD) / (pTD * pD + pTnotD * (1 - pD))
(pTD * pD_T1) / (pTD * pD_T1 + pTnotD * (1 - pD_T1))
```

```
## [1] 0.495
```

With three positive tests, it is almost certaint that the person has the disease,

```
(pTD^3 * pD) / (pTD^3 * pD + pTnotD^3 * (1 - pD))
```

```
## [1] 0.9898
```

If due to family history, the person had a 25% probability of having the disease prior to the test, after a positive test, it is almost certain that that person has the disease.
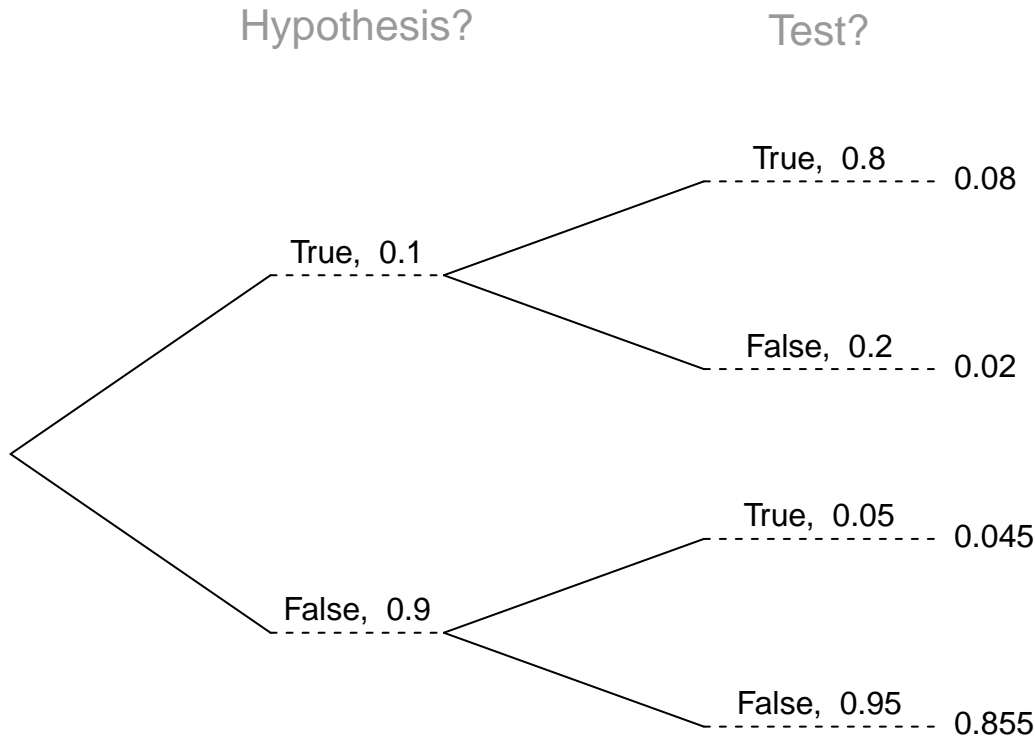
```
pD <- 0.25
(pTD * pD) / (pTD * pD + pTnotD * (1 - pD))
```

```
## [1] 0.9705882
```

Testing scientific theories

Suppose the probability the probability that a scientific theory (hypothesis) is correct is 10%. Afterall, we're on the edge of knowledge. The probability that a study confirms a hypothesis when the hypothesis is false is 0.05 (false positive or type I error). This is the standard accepted level of statistical significance. The probability that a study rejects a hypothesis when the hypothesis is false is 0.2 (false negative or type II error). What is the probability that a hypothesis is true given a study confirms it? Calculate this using a tree and Bayes' Theorem.

```
treeDiag(c("Hypothesis?", "Test?"),
         p1 = c(0.1, 0.9),
         p2 = list(c(0.8, 0.2),
                   c(0.05, 0.95)),
         out1 = c("True", "False"),
         out2 = c("True", "False"))
```

Hypothesis?   Test?

True, 0.1

True, 0.8 --- 0.08

False, 0.2 --- 0.02

False, 0.9

True, 0.05 --- 0.045

False, 0.95 --- 0.855

Let H be the event that the hypothesis (theory) is true, and $\neg H$ be that the hypothesis is not true. Let D ($\neg D$) be the event that the study supports (does not support) the hypothesis. We are interested in finding $\Pr(H|D)$, the probability that a hypothesis is true given that a study supports it. By Bayes' theorem,

$$
\begin{aligned}
\Pr(H|D) &= \frac{\Pr(D|H)\Pr(H)}{\Pr(D)} \\
&= \frac{\Pr(D|H)\Pr(H)}{\Pr(D|H)\Pr(H) + \Pr(D|H)\Pr(\neg H)} \\
&= \frac{0.8 \times 0.1}{0.8 \times 0.1 + 0.05 \times 0.9} \\
&= \frac{80}{80 + 45} \approx 0.64
\end{aligned}
$$

So given a study supports a hypothesis, there is only a 64% probability that the hypothesis is true. However, we should not be too pessimistic about this result. First, 64% is much higher than our prior probability that the theory was correct: 10%. This also implies that the probability that a hypothesis is true given that a study does not support it is very low,

$$
\begin{aligned}
\Pr(H|\neg D) &= \frac{\Pr(\neg D|H)\Pr(H)}{\Pr(\neg D)} && [t]\\
&= \frac{\Pr(\neg D|H)\Pr(H)}{\Pr(\neg D|H)\Pr(H) + \Pr(\neg D|\neg H)\Pr(\neg H)} \\
&= \frac{0.2 * 0.1}{0.2 \times 0.1 + 0.95 \times 0.9} \\
&= \frac{20}{20 + 855} \approx 0.02
\end{aligned}
$$

The important thing to note is that $P(H|D)$ is not the $1 - p - value = 0.95$. If we use a 0.05% p-value as our standard, the probability that the hypothesis is true is not 95%. The p-value is only the probability of a false positive. To calculate the probability that the hypothesis is true you also need the power of the test ($P(D|H)$) and the prior probability of the hypothesis ($P(H)$), which is generally unknown.

What happens to the probablity that the hypothesis if we increase our standard of evidence by lowering the Type I errors to 0.01 while keeping the power the same at 0.8?

$$
\begin{aligned}
\Pr(H|D) &= \frac{\Pr(D|H)\Pr(H)}{\Pr(D)} \\
&= \frac{\Pr(D|H)\Pr(H)}{\Pr(D|H)\Pr(H) + \Pr(D|H)\Pr(\neg H)} \\
&= \frac{0.8 \times 0.1}{0.8 \times 0.1 + 0.01 \times 0.9} \\
&= \frac{80}{80 + 9} \approx 0.89
\end{aligned}
$$

The probability that the hypothesis is correct is now 89%.

What happens to the probablity that the hypothesis is true if we perform another independent study? Let D' ($\neg D'$) be the event that the study supports (does not support) the hypothesis. Instead of using $P(H) = 0.1$, we can use our original result and use $P(H|D) = 0.64$.

$$
\begin{aligned}
\Pr(H|D') &= \frac{\Pr(D'|H)\Pr(H|D)}{\Pr(D'|H)\Pr(H|D) + \Pr(D'|\neg H)\Pr(\neg H|D)} \\
&= \frac{0.8 \times 0.64}{0.8 \times 0.64 + 0.05 \times 0.37} \\
&= \frac{512}{512 + 18.5} \approx 0.96
\end{aligned}
$$

This is even more effective than lowering the probability of a type 1 error. However, a key assumption of this calculation is that the studies were independent, so that $P(D\&D'|H) = \Pr(D|H)\Pr(D'|H)$. If they are not independent, it is likely that $P(D\&D'|H) < \Pr(D|H)\Pr(D'|H)$ since similar studies would likely come to similar conclusions. This suggests that the most valuable evidence for a theory are independent studies, e.g. those using different data or different methods to test a hypothesis.